31-10-2016

Deliverable DNA3.5: Recommendations and template policies for the processing of personal data

Deliverable DNA3.5

Contractual Date:	31-10-2016
Actual Date:	31-10-2016
Grant Agreement No.:	653965
Work Package:	NA3
Task Item:	TNA3.5
Lead Partner:	KIT
Document Code:	DNA3.5
Authors:	Uros Stevanovic, Ian Nielson, David Kelsey, Gerben Venekamp, Stefan Paetow, David Groep, Mikael Linden, Marcus Hardt, Hannah Short, Peter Gietz, Rob van der Wal

The research leading to these results has received funding from the European Community's Horizon2020 Programme under Grant Agreement No. 653965 (AARC).

Abstract

Collaboration among scientist across different administrative domains and across borders in Europe and beyond needs to address the management of personal information. In the majority of use cases, and for all cross-domain resource providers and e-infrastructures, providers will process personal data in order to provide services. The main objective is to provide templates for policies that operators of different components in the infrastructures can use. It provides updates about the legal context and identifies the minimal set of information needed by the participants in the prevalent use cases and those foreseen for the proof-of-concepts in AARC SA1.

Table of Contents

Executive Summary

Collaborations between scientists spanning administrative domains and borders in Europe and beyond are served by compute and storage services offered within research and e-Infrastructures. As a result, both must store and process the personal information of their users. In the majority of use cases, and for all cross-domain resource providers and e-infrastructures, providers will process personal data in order to measure usage and allocation of resources. To preserve the privacy of the individuals involved this personal data needs to be suitably protected whilst, at the same time, being made available to those who need it in order to be able to offer a service. Consequently, in designing or managing services, operators of these infrastructures must take decisions with potential legal consequences, with which are they not necessarily familiar with, potentially making scientific cooperation more difficult and thereby less attractive.

The overall goal of this task is to provide templates for policies that operators of different components in the infrastructures can use. This task started with the collection of use case studies of the processing of personal data from user communities and infrastructures (as suppliers or associations) in its Milestone[1] in Month 7. This deliverable extends this work, and provides updates about the legal context as well as proposing template policies for the infrastructure components defined in the Blueprint Architectures in JRA1. Policies for attribute release by Identity Providers (IdP) to Service Providers (SP) or the processing of user-provided data, that may contain Personal Identifiable Information (PII), are not covered. These are complex issues beyond the scope of this work. Recommendations are provided for policies that can be applied across an entire infrastructure – keeping in mind the current state of European legislation, specifically the new General Data Protection Regulation (GDPR)[2] and its implementation in the Member States. It identifies the minimal set of information needed by the participants in the prevalent use cases and those foreseen for the proof-of-concepts in AARC SA1.

[1] https://aarc-project.eu/wp-content/uploads/2015/11/MNA3.2-AccountingDataProt-20151130.pdf

[2] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

1 Introduction

The highly collaborative nature of today's research has resulted in the establishment of both domain-specific Research Infrastructures ('RIs') and more generic 'e-Infrastructures' that serve multiple research domains. Both of these infrastructure types (hereafter commonly referred to as 'Infrastructure'), by their very nature, span multiple organizations (i.e. administrative domains) and are often transnational in character. Many have global scope, incorporating researchers and services from around the world. Infrastructures act as providers of coherently organised services, provided by and under the authority of the independent organizations that jointly make up the Infrastructure. Yet, these independent organizations must work together to provide collective services, offered to their common users as if they were a single service. Moreover, in many cases, Infrastructures work together and have to exchange data between themselves.

In order to provide their services, Infrastructures have a need to exchange personal data, *e.g.*, to measure actual service use versus allocated resources; to satisfy regulatory requirements demanded by funding agencies or European Commission grant agreement rules[3]; to comply with export controls[4]; or to distribute resources fairly amongst their users. Because of the inherently distributed nature of Infrastructure AAI's (Authentication and Authorization Infrastructures), the data collected by Infrastructures includes information from multiple sources, data that may contain personally identifiable information (PII) about the user. This data needs to be shared between the ensemble of service providers in the Infrastructure spanning administrative domains, across borders in Europe, and beyond at the global level. To preserve the rights of the user of the Infrastructure (the researcher), and ensure that the organizations providing the Infrastructure comply with applicable national and EU regulatory requirements, the collection and sharing of PII must be managed.

The purpose of this document is to provide recommendations and template policies to resource providers and user-communities that establish and operate Infrastructures. These recommendations intend to facilitate the ability to collect, transfer, provide access to, and/or publish data related to the accounting, monitoring, logging, or any kind of processing of personal user data needed for the operation of the services provided by the resource providers.

This document presents:

- the scope (e.g. categories of data, reasons for data processing) listing the relevant data collected by the Infrastructure, to assist in identifying the applicable regulatory and legal frameworks on which decisions on how to handle this data should be based;

[3] http://ec.europa.eu/research/participants/portal4/desktop/en/funding/reference_docs.html#h2020-mga-gga

[4] http://www.wassenaar.org/

- a review of the relevant elements of existing legal privacy frameworks, primarily from European and, if applicable, from national bodies, with specific regard to the changes resulting from the adoption of the new EU General Data Protection Regulation (GDPR) in April 2016;
- a discussion of the policies of communities and infrastructure providers regarding collection and processing of personal data that provide potential models for data sharing;
- specific recommendations and suggestions on how to leverage a policy framework to permit sharing of personal data within a coherently-organised Infrastructure.

This work build upon results previously presented in *Requirements on data to protect from AAI, community, resource providers and e-infrastructure* (published as AARC document MNA3.2). New use cases identified since that document was written, including those identified through the AARC SA1 pilots, have been added as background information.

Although recommendations in the present document are made with the intention to align with applicable EU directives and regulations, and the conceptual framework presented here has been discussed with experts in the field, the specific recommendations and arguments presented in this document must not to be considered as legal advice in any particular jurisdiction.

2 Scope

User communities, resource providers, research communities, Research Infrastructures (RI), e-Infrastructures, as well as specialized services (e.g. IdP-SP proxies[5]) have a need for collecting and processing personally identifiable information. Such personal data are any kind of information that can be used to identify an individual.

When we speak about the processing of the personal data, we will focus on the following points:

- Collection of usage data in RI and e-Infrastructures
- Correlating resource usage to people and groups
- Accumulation of usage data across countries (and continents)
- Collection and processing of personal data for incident response

The scope of the document relates to personal data processing necessary for accounting, monitoring, and collaboration. Within this work, we will NOT provide information or policies covering the usage and handling of research data sets that may contain personal information (e.g. medical data). Furthermore, the question of user attribute release (i.e. an IdP providing data to an SP) in research communities and federated environments is not within the scope of this document. Policies and procedures related to the attribute release by an Attribute Authority (e.g. DP CoCo[6] as managed by REFEDS[7]) depend on the outcome of other on-going related architecture work, and such issues will have to be addressed by future work - at the time of writing foreseen for 2017-2018. However, if attributes labelled as personal identifiable information are used for the same reasons as in the scope of this document, then the same policies for processing should be applied as for other personal data discussed here.

We also assume that all activities undertaken on the Infrastructures are 'professional' work, and anyone interacting with the Infrastructures does so while appropriately endorsed by a real 'legal entity', e.g., a researcher is employed by an organization (university, research laboratory, &c) and thus (maybe implicitly) also binds the organization in anything she or he does. This is particularly relevant for user community administrators who have access to usage data and similar handling of user's data.

Primarily, we are focusing on the requirements for research conducted within the European communities. However, as much of the research has a substantial international element, when making the recommendations we take into account the practice and policies for the accounting of personal data that exists outside of Europe.

[5] https://spaces.internet2.edu/display/GS/SAMLIdPProxy

[6] http://geant3plus.archive.geant.net//uri/dataprotection-code-of-conduct/v1/Pages/default.aspx

[7] https://refeds.org/

3 Background and recent developments

In the report *Requirements on data to protect from AAI, community, resource providers and e-infrastructure* (published as AARC document MNA3.2), a representative selection of Infrastructures was analysed and the roles of the various participants in the Infrastructure assessed in view of national and European regulatory information. This work serves as the starting point for the current *Recommendations*, bearing in mind that the *Requirements* were published before the new General Data Protection Regulation (GDPR) had been adopted, and thus reflect mainly the then-current Data Protection Directive 94/46/EC[8] and its implementation in national law within the EU. Although the GDPR differs from the earlier Directive, it retains the same elements as the basis for protecting personal data and, as such, the requirements identified in MNA3.2 need only to be re-assessed.

The *MNA3.2 milestone* provides an overview of the (representative) sample of existing communities and infrastructure provider policies regarding the processing of personal data, and Infrastructures such as EGI and PRACE are discussed therein. Community requirements on accounting data have common, well-established policy sets, evolved over the last decade or more. As such, participants in these infrastructures, including both providers and consumers of compute, storage or other services, are required to comply with the rules and procedures laid out in the relevant policies. This includes not only the handling of accounting data discussed in the *Requirements*, but also other operational and security areas such as the registration of users, their experiment affiliations, retention of service logs and security incident response data. All of these are necessary for the efficient operation and management of an infrastructure, and many require the storage and processing of Personal Data. Naturally, not all communities and resource providers were considered, however, an effort was spent to review sufficiently many policies to make valid general conclusions. Two cases, based on the pilots conducted in the AARC SA1 activity, are added at this stage as they include new elements – Umbrella ID devolving all data processing responsibilities to its constituent organizations and the generic case of 'cloud' services proposing a 'Code of Conduct' approach.

In addition to the above, Infrastructures commonly serve communities that are truly global in their extent, and as such must, in the way they store and process Personal Data, take this global scope into account. Both the EC Directive and the new GDPR allow for such International transfers under a limited number of circumstances (derogations or exemptions). Their most relevant data protection aspects are discussed in the following sections.

[8] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

3.1 Umbrella ID

Umbrella ID[9] is an identity system designed by the European photon and neutron source facilities (PaNs). It aims to make life easier and science more productive for both the facilities and their users. Umbrella provides any PaN-user (and effectively anyone interested in scientific discovery) with a unique identity, the Umbrella ID. Equipped with such an ID a user can virtually access participating facilities with a single sign-on. Since the same Identity is known at each of the facilities, a user can more simply access or share data, manage administrative processes or make use of services and infrastructures provided by the PaNs. Umbrella is a joint project of the PaNs and other facilities with similar needs for an Identity Management System, and currently operates under a Memorandum of Understanding (MoU) signed by all participants in the collaboration.

Currently the MoU states the following in respect to accounting and processing of personal data:

> 5.6 A Party providing a service has the responsibility and rights for that service. All services shall be made available to all users, but are subject to authorization by the Party offering the service. For example, any user may implicitly be authorized on an open access scientific database or a software catalogue. On the other hand, the WUOs will always demand a local registration and, in some cases, certain documents like a passport before granting access to beam lines or facilities.

Therefore, the onus on processing and handling of personal data falls to the local user office (WUO), e.g. the facility, and facilities have amended their terms and conditions of use accordingly. The only attributes shared between the Umbrella ID system and any local facility are a user ID (the Umbrella ID itself) and two unique but non-personally identifiable UUID strings.

Developments by the Umbrella collaboration in the near future include an attribute authority to allow the storage of additional attributes to be released to non-Umbrella collaboration services. In this instance, the collaboration expects to adhere to the existing GÉANT Data Protection Code of Conduct for self-asserted attributes with support for the standard Shibboleth 3 attribute release interstitial flow during login.

3.2 Cloud Computing

In recent years, there has been widespread adoption of the Cloud Computing[10] model for the provision of commercial compute services. Similarly, there are a number of projects trialling its use in the research sector. For example, Helix Nebula[11], “Europe’s Leading Public-Private Partnership for Cloud” and the Indigo

[9] http://info.umbrellaid.org/index.php?id=2

[10] https://en.wikipedia.org/wiki/Cloud_computing

[11] http://www.helix-nebula.eu/

DataCloud[12] aimed at "developing a data and computing platform … provisioned over hybrid (private or public) e-infrastructures".

For many years, research communities have made use of resources distributed across the Internet[13] and, whilst the technology and model are changing, many of the concerns relating to the protection of users' privacy arising from the exchange of operational data across the infrastructure remain the same. Whether the cloud-based resources used (for compute or data storage) are located within a "private cloud"[14] (resources dedicated to a single client) or "public cloud"[15] (resources shared in a dynamic fashion by multiple clients), responsibility, and subsequent liability, for ensuring that the cloud provider fulfils the necessary legal obligations remains with the client, just as in traditional "outsourcing" models. What does change is the level of difficulty in ensuring compliance. This is particularly so in present commercial settings where there may be multiple layers of dynamic service provision, involving several providers. For example, a provider for web-based email may provision all or part of their service on one or more virtualised cloud infrastructures, which may be transparently (from the point of view of both service user and provider) distributed across multiple data-centres in multiple countries or continents. The research sector is not at this level of cloud adoption yet, but the difficulty of the client organization (as "data controller" in EU terminology) in applying due diligence may be a significant obstacle to adoption in the future (see Millard et al[16] for a full discussion). Initiatives, encouraged in the General Data Protection Regulation, for the drawing up of a Code of Conduct for Cloud Service Providers[17] are intended to help both the cloud providers (processors) and their clients in this process but remain a significant "work-in-progress". For the time being, existing policies adopted by Research Infrastructures may be sufficient to cover proposed cloud usage. However, cloud usage is an area that will need further consideration as adoption and complexity of the resulting infrastructures increases.

3.3 Changes to EU data protection rules and the new Regulation

On 8th April 2016, the European Council adopted the Regulation 2016/679, also known as the General Data Protection Regulation (GDPR). Directive (2016/680), related to the prevention, investigation, detection or

[12] https://www.indigo-datacloud.eu/

[13] http://wlcg.web.cern.ch/

[14] https://en.wikipedia.org/wiki/Cloud_computing#Private_cloud

[15] https://en.wikipedia.org/wiki/Cloud_computing#Public_cloud

[16] Cloud Computing Law – Millard et al. OUP 2013

[17] https://ec.europa.eu/digital-single-market/en/news/data-protection-code-conduct-cloud-service-providers

prosecution of criminal offences with regard to the processing of personal data, was also adopted. However, within the scope of this document the GDPR is of more consequence. This European Regulation establishes the legal framework for the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and it repeals the Directive 95/46/EC that is currently valid. It went into force on the 24th May 2016, but it will apply from the 25th May 2018[18], when the current Directive will be repealed. Being a Regulation, the GDPR is legally binding for all Member States, without the need to be ratified by Member State parliaments. It is a key document used as a basis for our discussion and recommendations.

[18] http://ec.europa.eu/justice/data-protection/reform/index_en.htm

4 Interpretation of the regulation in the context of shared infrastructures

As the GDPR comes into effect, it is important to address definitions and main points coming from the GDPR. This chapter covers those points, and explains them in simpler terms.

4.1 Personal data and processing

Here we present how the GDPR defines personal data, and what it means to process personal data. These definitions are quoted from GDPR, and explained, since they are used throughout this document.

Personal data (Article 4(1)):

> "Any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."

In other words, *personal data* is any data set that can be used or combined, from and with, any source that can be used to determine information about a natural person.

Processing (Article 4(2)):

> "Any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction."

In summary, any operation on personal data is considered processing, and as such, needs to be managed (e.g. by policies).

4.2 Data Processors and Data Controllers

Article 4.7 of the GDPR defines the role of *data controller* as someone (natural or a legal person, one acting in the capacity of a data controller) who decides how and in which ways personal data will be processed. Article 4.8 of the GDPR defines the *processor* as someone (again, either a natural or a legal person) who processes the data on behalf of the data controller.

In short[19], a data controller is "a person who (either alone or jointly or in common with other persons) determines the purposes for which and the manner in which any personal data are, or are to be, processed." A data controller is the responsible party that must ensure that all processing of personal data complies with the GDPR. Failure to do so may result in legal repercussions. Data processors on the other hand process personal data solely under the direction of a data controller who decides what personal information will be kept and to what uses it may be put.

There are eight rules of data protection that each data controller must ensure are followed[20]:

- Personal Data must be processed legally and fairly
- It must be collected for explicit and legitimate purposes and used accordingly
- It must be adequate, relevant and not excessive in relation to the purposes for which it is collected and/or further processed
- It must be accurate, and updated where necessary
- Data controllers must ensure that data subjects can rectify, remove or block incorrect data about themselves
- Data that identifies individuals (personal data) must not be kept any longer than strictly necessary
- Data controllers must protect personal data against accidental or unlawful destruction, loss, alteration and disclosure, particularly when processing involves data transmission over networks. They shall implement the appropriate security measures
- These protection measures must ensure a level of protection appropriate to the data

We found that, in our use cases, where the collaboration consists of independent cooperating organizations, data controller is the role we encounter most of the time. A data processor role is separated from the controller in few situations. Furthermore, it is also common to have *Joint controllers* within the use cases we are describing. From the GDPR Article 26.1, "where two or more controllers jointly determine the purposes and means of processing, they shall be joint controllers". In this case, they should define their respective legal responsibilities, and users have the right to address their rights with each of the controllers.

4.2.1 Examples

The GDPR gives the definitions of data controller, data processor, etc. What does this mean in real life? What are the roles one may encounter stated in the GDPR? For example, a Virtual Organization[21] (VO) manager decides which people have access rights to the VO he/she manages. They does so based on identities, which makes the organization to which VO manager belongs a data controller. A VO manager could appoint (or

[19] https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/

[20] http://ec.europa.eu/justice/data-protection/data-collection/obligations/index_en.htm

[21] https://www.opensciencegrid.org/about/organization/

actually designate, since the VO usually has no legal organizational structure) one or more persons (natural or legal) to act on their behalf, i.e. deciding how and which data is processed. Those persons are therefore joint controllers according to the GDPR, i.e. they jointly decide how and for which reasons the data are processed.

IdPs, in the form of authorization attributes, provide identity information to SPs. This means that the organization (or organizations, in which case they are joint controllers) to which IdP belongs is responsible for assembling/defining personal data is the data controller. It/they must ensure that the processing activities are compliant with EU data protection law. In collaborative research scenarios, in most cases the SPs are also data controllers. The SPs in rare cases may be data processors on behalf of the IdP, e.g. SP is an IaaS (Infrastructure as a Service) provider.

Federations come in two flavours: hub-and-spoke and mesh. In mesh federations, there is no intermediary between an IdP and SP. With hub-and-spoke federations, there is an extra entity between an IdP and SP. This entity acts as a proxy for the federation and is therefore both an IdP and an SP, and acts as a data controller (if these store and augment information) and data processor.

4.3 Purpose of processing

The controller defines the purpose of processing of personal data (Article 5 b).

It can include:

- ensuring the integrity, availability and confidentiality of the infrastructure (i.e. information security)
- monitoring the resource consumption and, if necessary, invoicing
- capacity planning

For example, EGI defines the purpose of processing in the EGI AUP[22]: as “You agree that logged information, including personal data provided by you for registration purposes, may be used for administrative, operational, accounting, monitoring and security purposes. You agree that this logged information may be disclosed to other authorised participants via secured mechanisms, only for the same purposes and only as far as necessary to provide the services.”

One reason for processing and collecting personal data for security purposes would be security incident response. As mentioned previously, the user should be informed of all the data processing use cases for which their data is used.

[22] https://wiki.egi.eu/wiki/SPG:Drafts:Acceptable_Use_Policy_March_2015

4.4 Legal grounds for data processing

The Regulation, as the Directive before it, recognises six distinct legitimate grounds for processing personal data. Of these, two are by nature not applicable to the research and collaborative use cases considered in the Infrastructures: vital interest of the data subject (since there is no life-threatening situation), and tasks carried out in the public interest or official authority (since either should be bestowed upon the controller by EU or member state law). There are also few reasons in which processing based on legal obligations is relevant, with the only likely case being the recording of nationality for compliance with the Wassenaar Arrangements. The nature of research collaboration for all practical purposes precludes "performance of contract" as a useful basis for processing, leaving only user consent (6.1a) and legitimate interest (6.1f) as applicable processing grounds.

However, GDPR places stringent requirements on user consent. The consent has to be given clearly and freely, only for a stated purpose, and can be revoked at any time. Users must agree to abide by the appropriate infrastructure policies (e.g. AUP) and, in doing so, are informed that their Personal Data will be shared within the infrastructure for stated, limited purposes. Such acceptance by the user might be considered as consenting to the transfer of data. However, as a researcher will, in many cases, be required to gain access to the services the infrastructure offers as a prerequisite to performing data analysis which is critical to their employment as a researcher, this implies that such consent could potentially not be seen as a "freely given, specific, informed and unambiguous indication of data subject's wishes" (Article 4.11).

Also, from Article 25 (privacy by design), data processing must include certain principles (such as data minimization) before asking for the user consent. Otherwise, the consent may not be legally binding. For example, this court decision mentions these points (in German)[23], summarized also here (in German)[24].

Previous work[25] conducted in the context of federated attribute release, which is relevant also to this case, indicates that the use of 'legitimate interests' of the controller continues to be a good basis for processing personal data in the Infrastructure. Furthermore, recital 49 of the GDPR recommends the same detail.

[23] https://www.ldi.nrw.de/mainmenu_Service/submenu_Entschliessungsarchiv/Inhalt/Beschluesse_Duesseldorfer_Kreis/Inhalt/2016/Fortgeltung_bisher_erteilter_Einwilligungen_unter_der_Datenschutz-_Grundverordnung/Fortgeltung_bisher_erteilter_Einwilligungen_unter_der_Datenschutz-_Grundverordnung1.pdf

[24] https://www.heise.de/newsticker/meldung/Datenschutzexperten-kritisieren-Ansichten-der-Datenschutz-Aufsicht-zur-Einwilligung-3329888.html

[25] Cormack, A., *Federated Access Management and the GDPR*, 4 February 2016, https://community.jisc.ac.uk/blogs/regulatory-developments/article/federated-access-management-and-gdpr, visited October 2016

4.5 Release of personal data to 3rd parties, data subject rights and retention

The use of legitimate interests of the data controller as a basis for processing and the structure of the Infrastructure makes almost all entities (including service providers) a controller in the Regulation sense. Any sharing of data within the Infrastructure – log files, accounting records, inferred community membership information – must be considered 'release' of personal data to a third party. Similarly, sharing such information with the Users' home organization (research institute, university) or with the people responsible in the community for resource allocation and fair use of the shared resources in the Infrastructure is release to a third party.

Disclosure to third parties is permitted when certain safeguards and controls are in place. The individual must be informed of the fact that sharing will take place for such legitimate interests, and the Opinion 06/2014 of the Article 29 DP Working Party[26] provides the criteria that make such processing legitimate. Although based on the Directive, since the differences between the "old" Article 7.f and Article 6.1f from the GDPR are minimal, we presume that the opinion is still valid. In the Opinion, it is stated that Article 6.1f should not be treated as last resort, nor be automatically applied. It provides a balancing test, which, in short, considers both legitimate interests of the data controller and data subject: the stronger the legitimate interest being pursued by the data controller and the less harm the processing does to the interests of the data subject, the greater the likelihood that the activity will be lawful[27]. For example, information disclosure about a user by the Identity Providers has a positive rather than negative impact on the user, in line with the user's expectation. However, one can argue the legitimate interest of the Service Providers to have information about the users in case of security incidents. Data minimisation and privacy enhancing technologies should be employed.

The user has a right (Articles 12-15) to be informed about the purpose of data collection and processing. It should be provided clearly, and at a moment where such data is collected (e.g. when the user starts using the infrastructure). It should inform the user who the data controller is, the means of contacting the data controller and also, how the user may request to rectify incorrect data (Article 16). One point to mention is Article 21.1, where the user may object to processing of data. Here, the controller can still continue the processing of data if he or she can show "compelling legitimate grounds" for doing so (e.g. storing log data may be a legitimate interest which overrides the user request for not storing the user data in log files). The duration of storing personal data is not explicitly defined in the GDPR, other than it should not be kept for longer than needed, e.g. for invoicing, legal compliance, incident response etc. In addition, appropriate technical security measures should be employed to keep the user personal data safe (Articles 24, 25, and 32).

[26] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf

[27] https://community.jisc.ac.uk/blogs/regulatory-developments/article/legitimate-interests-and-federated-access-management

4.6 Release of personal data outside the European Union

Almost all Infrastructures are global in nature, or at the very least contain participants (both users as well as service providers) outside the European Union. Release of personal data to independent data controllers outside the European jurisdiction requires the data exporter to ensure the adequacy of protection.

Under the Directive, the European Commission has so far recognized just a few countries: Andorra, Argentina, Canada (commercial organizations), Faeroe Islands, Guernsey, Israel, Isle of Man, Jersey, New Zealand, Switzerland and Uruguay that are deemed to provide adequate protection. In addition, the new Regulation explicitly mentions release to International Organizations and, as such, any such organizations participating in the Infrastructure need to be assessed as well. Since quite a few Infrastructures include organizations that have been established by treaty (e.g. many of the EIROForum members such as CERN, EMBL), they also require such non-EU treatment until their data protection has been assessed by the Commission – until now these organizations provided an open place to exchange personal data and serve the users and individuals without such considerations.

For transfer of personal data internationally, a number of conditions (appropriate safeguards) are defined permitting such transfer to take place, and the user must be informed of these safeguards. Apart from explicitly obtaining approval for each exchange model by a data protection authority – which with the scale of research collaboration is not a viable proposition – only a few viable mechanisms to provide for safeguards remain: 'binding corporate rules', 'standard data protection clauses', and 'approved Codes of Conduct'. Explicit user consent is always an option, yet the requirements on user consent (as discussed also for the intra-European case) do not make it a suitable ground for processing within the Infrastructures. Codes of Conduct need to be so specific in order to be approved – this is the conclusion from the discussions with members of WP29[28] following the ongoing work on Codes of Conduct in the context of REFEDS and GEANT . Significant effort in this area is needed to match it to the collaborative nature of the Infrastructures.

For a long time, additional mechanisms have been in place to permit the exchange of personal data with organizations based in the United States of America. Initially through the "Safe Harbour" model (found invalid by the European Court of Justice in 2015), and currently through the new EU-U.S. Privacy Shield[29], administered in the US by the Department of Commerce[30] and adopted by the European Commission. The arrangement includes:

- strong data protection obligations on companies receiving personal data from the EU
- safeguards on U.S. government access to data;
- effective protection and redress for individuals;

[28] https://wiki.refeds.org/display/CODE/Data+Protection+Code+of+Conduct+Home

[29] http://ec.europa.eu/justice/data-protection/international-transfers/eu-us-privacy-shield/index_en.htm

[30] http://ec.europa.eu/justice/data-protection/article-29/press-material/press-release/art29_press_material/2016/20160726_wp29_wp_statement_eu_us_privacy_shield_en.pdf

- annual joint review to monitor the implementation.

Neither Safe Harbour nor the EU-US Privacy Shield is of any use to the Infrastructures, since it is exclusively limited to commercial entities. As such, it is too limited in scope to be applicable to the general Infrastructure case, where research labs and universities mostly collaborate.

In practice, the Regulation framework and the necessity to exchange personal data globally in the Infrastructures, leave only the Standard Data Protection clauses ("Model Contracts") and Binding Corporate Rules ("BCR") as viable options on which to base personal data sharing in the Infrastructures. However, we must acknowledge that "BCRs" have to be approved by a competent supervisory authority and must be legally binding – both of which prerequisites are impractical for the use cases considered here.

Yet, the methodology behind the BCRs gives practical guidance that in view of the risk exposure could be considered a suitable operating model. This is particularly true considering that the risk of harm to the user by exposure of their data is considered very low. The accounting data is already being open to consultation within the EU by persons (organizations) that can demonstrate a legitimate interest – and the same conditions imposed on these persons within the EU are also imposed on those outside the Union (following the reasoning behind article 49 (f)). The policy set of the Infrastructure in both cases provides for the 'balancing test' necessary between legitimate interest and rights and freedoms of the data subject.

5 Recommendations

The new General Data Protection Regulation has the headline-grabbing change of greatly increasing the maximum penalties for non-compliance – 20 million Euro or 4% of worldwide turnover. In increasing the risk to business of failing to ensure appropriate measures are in place to protect citizens privacy, this change may also have some effect of stifling the willingness of research organizations to collaborate due to both fear of consequences of (accidental) non-compliance and the weight and length of the process which must be gone through to comply. In practice, considering the requirements on cross-border transfers, there are significant changes to the detail of the legislation, such as requirements on Data Protection Officers, but the overall framework remains. Both Standard Data Protection Clauses (Model Contracts) and Binding Corporate Rules (BCR) benefit from changes related to consistent application across the Member States.

One significant addition to the GDPR related to international transfer is the allowance that such transfers will be able to occur where there is an approved Code of Conduct in place. Such Codes of Conduct are intended to target organizations engaged in common industry sectors where adherence may be taken as evidence of compliance to GDPR. One example of such a prototype Code of Conduct is that being developed for Cloud Service Providers by the Cloud Select Industry Group (C-SIG)[31]. One could envisage that a Code of Conduct for research Infrastructure services might be considered in the future as a standard component of an Infrastructure policy set.

Based on the discussion of the Regulation and global nature of Infrastructures, only two viable models remain, i.e. model contracts and binding corporate rules. We review each of these for suitability.

5.1 Standard Data Protection Clauses (Model Contracts)

Two standard forms of contract, referred to as Model Contracts[32] (MC), have been issued by the EU covering the cross-border, controller-to-controller or controller-to-processor, transfer of personal data. Where each party to a transfer is able to sign such an agreement, the controller or processor located outside of the EEA is deemed to offer adequate protection to users' personal data as a destination for a transfer.

Yet the model clauses must be part of a contractual agreement, which presumes such a contract is in place between the parties transferring personal data. The Infrastructures, being composed of a large number of independent organizations, are not usually based on contracts, and most certainly do not use bilateral contracts between all participants – the combinatorics do not permit such to happen. Case studies also indicate that, unless a specific agreement is in place, contracts 'by proxy' (one organization including the model clauses in a

[31] https://ec.europa.eu/digital-single-market/en/cloud-select-industry-group-code-conduct

[32] http://ec.europa.eu/justice/data-protection/international-transfers/transfer/index_en.htm

contract, and any organization loosely affiliated with the signing organization being covered implicitly) is also not acceptable as a formal legal basis.

Yet model clauses in specific contexts are a suitable mechanism: their applicability to the procurement of commercial 'cloud' services has been demonstrated (as part of the GEANT project as well as in specific 'normative reference frameworks' that have been adopted by national research network organizations on behalf of their constituents in 'joint procurements').

Where the structure and pre-existing legal framework exists, model clauses could be a workable solution to the data transfer problem – their use is standard and their application need not be further elaborated here. For research and collaborative infrastructures that span globally, these are however not useful.

5.2 Binding Corporate Rules

Whereas Model Contracts, mentioned above, are fixed texts, Binding Corporate Rules (BCR) are drafted by the organization itself and, as such, can be worded to the context and environment within which they operate. Once approved by the appropriate Data Protection Authority, BCRs permit legal transfer of data, internationally, within the body bound by the BCR. The EU has a short overview on BCR and the accompanying procedures[33], and WP29[34] has issued detailed guidance on the creation of BCRs.

Binding Corporate Rules are used to provide for sufficient insurance for the protection of privacy of individuals as mentioned in article 26 (2) of the Directive 95/46/CE for all transfers of personal data protected under a European law. The purpose of BCR is that personal information transfers are adequately protected, thereby negating the necessity to sign individual contracts between parties within the group. The BCR are especially useful for the Infrastructure cases where the personal information is transferred outside of the European Economic Area and where the data protection is not sufficient according to EU Law.

BCRs contain at least the following elements:

- Privacy principle including transparency, data quality, security of the data, etc.
- Tooling to measure the effectiveness, for example audits, training, complaint handling, etc.
- Clear statement that the BCR are binding.

[33] http://ec.europa.eu/justice/data-protection/international-transfers/binding-corporate-rules/index_en.htm

[34] http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2015/wp204.rev_en.pdf

5.2.1 Adopting Binding Corporate Rules

In order to adopt BCR, a formal procedure needs to be followed and involves the national Data Protection Authorities (DPA). The various elements of the BCR need to be reviewed by the national DPA to ensure that it meets the criteria as set out by the Article 29 Working Party. The procedure recognizes one lead authority. This means that the company can conduct the communications on behalf of all individuals and they do not have to be approached separately. Next to the lead authority is the mutual recognition group. This group exists to speed up the EU procedure. Some DPAs are members of this group and currently these are *Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Estonia, France, Germany, Iceland, Ireland, Italy, Latvia, Liechtenstein, Luxembourg, Malta, the Netherlands, Norway, Slovakia, Slovenia, Spain, and the United Kingdom*.

In order to get approval for the BCR, the following five steps must be taken:

1. Determine and assign the lead authority that will handle the EU co-operation procedure amongst the other European DPAs;
2. The company will draft the BCR and therein address Article 29 Working Party. The lead authority will read and comment the document to ensure the document is in agreement with Working Party 153;
3. The lead authority will consult the relevant DPAs and thereby start the EU cooperation procedure. Relevant DPAs are of the countries where the personal information is exported to and do not have sufficient data protection.
4. Closing of the EU cooperation procedure happens when:
    a. the countries of the mutual recognition acknowledged the receipt of the BCR;
    b. DPAs not under mutual recognition have verified that the BCR is in accordance with the requirements as set out in Article 29 WP;
    c. responses need to be sent within one month.
5. Once the BCR have been considered as final by all DPA, the company shall request authorisation of transfers based on the adopted BCR by each national DPA.

The approach currently in process of adoption by the EGI and WLCG Infrastructures follows the model of and guidance concerning Binding Corporate Rules in the context of an e-Infrastructure bound by an existing policy set. The infrastructure requires all participants[35] to adhere to an eight-point policy for the processing of personal data that aims to follow the WP29 guidance mentioned above. This provides a relatively lightweight framework that is seen as assisting its adoption by service providers and user communities. Of course, the participants in the Infrastructure do not form a legal corporation and as such, this approach could be seen as inapplicable, not least because the assignment of risk between the participants is not governed by legally enforceable agreements. However, in mitigation of this, the risk of harm to the user by exposure of their data in contravention of the policy is considered very low, due to both the lack of sensitivity of the data being processed and the fact that in many circumstances the data is already available from other sources such as research publications etc.

[35] As already described in the scope of this document, individuals who are participants in the Infrastructure are always (at least officially) acting *on behalf of* their organization – and this does not include people acting on a purely personal basis.

5.3 Conclusion and Recommendation

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 signed form is then lodged with the federation. This approach, still a work-in-progress, remains a relatively complex, somewhat lengthy legal document, which may hinder adoption.

In practice, the AAI architecture proposed by the AARC project could in fact utilise both MC and BCR. The use of a Proxy Identity Provider, which acts to Infrastructure services both as a single source of identity information, from users' home organizations, and an aggregator of users' experiment or research affiliations, typically their VO membership and roles, may act as a pivot point between the two domains. On one side, the Proxy IdP participates as a service to users' home IdPs, caching the resulting attributes and executing the necessary federation agreements, possibly based on the MCs. On the other side, the Proxy IdP participates in the Infrastructure, bound by the BCR-like policy framework, acting as a source of identity information to infrastructure services.

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*.

6 Appendices

Appendix A Template *Policy on the Processing of Personal Data*

The following is the relevant text of the *Policy on the Processing of Personal Data*, developed jointly with EGI, WLCG, and the UK GridPP project. Being based on the principles of Binding Corporate Rules, it implements this recommendation guidance.

A.1 Introduction

This policy ensures that data collected as a result of the use of the Infrastructure is processed fairly and lawfully by Infrastructure participants. Some of this data, for example that relating to user registration, monitoring and accounting contains “personal data” as defined by the European Union (EU) [A.6 R1]. The collection and processing of personal data is subject to restrictions aimed at protecting the privacy of individuals.

A.2 Definitions

Infrastructure. The bounded collection of universities, laboratories, institutions or similar entities, which adhere to a common set of policies [A.6 R2] and together offer data processing and data storage services to End Users.

Participant. Any entity providing, managing, operating, supporting or coordinating one or more Infrastructure service(s).

Personal Data. Any information relating to an identified or identifiable natural person [A.6 R1].

Processing (Processed). Any operation or set of operations, including collection and storage, which is performed upon Personal Data [A.6 R1].

End User. An individual who by virtue of their membership of a recognised research community is authorized to use Infrastructure services.

A.3 Scope

This policy covers Personal Data that is Processed as a prerequisite for or as a result of an End User's use of Infrastructure services. Examples of such Personal Data include registration information, credential identifiers and usage, accounting, security and monitoring records.

This policy does not cover Personal Data relating to third parties included in datasets provided by the End User or the research community to which they belong as part of their research activity. Examples of such data are medical datasets that may contain Personal Data.

A.4 Policy

By their activity in the Infrastructure, Participants:

a) Declare that they have read, understood and will abide by the Principles of Personal Data Processing as set out below.
b) Declare their acknowledgment that failure to abide by these Principles may result in exclusion from the Infrastructure, and that if such failure is thought to be the result of an unlawful act or results in unlawful information disclosure; they may be reported to the relevant legal authorities.

A.5 Principles of Personal Data Processing

1. The End User whose Personal Data is being Processed shall be treated fairly and in an open and transparent manner.
2. Personal Data of End Users (hereinafter "Personal Data") shall be Processed only for those administrative, operational, accounting, monitoring and security purposes that are necessary for the safe and reliable operation of Infrastructure services, without prejudice to the End Users' rights under the relevant laws.
3. Processing of Personal Data shall be adequate, relevant and not excessive in relation to the purposes for which they are Processed.
4. Personal Data shall be accurate and, where necessary, kept up to date. Where Personal Data are found to be inaccurate or incomplete, having regard to the purposes for which they are Processed, they shall be rectified or purged.
5. Personal Data Processed for the purposes listed under paragraph ii above shall not be kept for longer than the period defined in a relevant Infrastructure service policy governing the type of Personal Data record being Processed (e.g. registration, monitoring or accounting) and by default shall be anonymised or purged after a period of 18 months.

6. Appropriate technical and organizational measures shall be taken against unauthorised disclosure or Processing of Personal Data and against accidental loss or destruction of, or damage to, Personal Data. As a minimum, Infrastructure Participants shall:
    a. Restrict access to stored Personal Data under their control to appropriate authorised individuals;
    b. Transmit Personal Data by network or other means in a manner to prevent disclosure to unauthorised individuals;
    c. Not disclose Personal Data unless in accordance with these Principals of Personal Data Processing;
    d. Appoint at least one Data Protection Officer (DPO) with appropriate training and publish to the Infrastructure a single contact point for the DPO to which End Users or other Infrastructure Participants can report suspected breaches of this policy;
    e. Respond to suspected breaches of this Policy promptly and effectively and take the appropriate action where a breach is found to have occurred;
    f. Perform periodic audits of compliance to this Policy and make available the results of such audits to other Infrastructure Participants upon their request.
7. Each Infrastructure service interface provided for the End User must provide, in a visible and accessible way, a Privacy Policy (see example policy in section A.7 below) containing the following elements:
    a. Name and contact details of the Participant Processing Personal Data;
    b. Description of Personal Data being Processed;
    c. Purpose or purposes of Processing of Personal Data;
    d. Explanation of the rights of the End User to:
        i. Obtain a copy of their Personal Data being stored by the Participant without undue delay;
        ii. Request that any Personal Data relating to them which is shown to be incomplete or inaccurate be rectified;
        iii. Request that on compelling legitimate grounds Processing of their Personal Data should cease;
    e. The contact details of the Participant's DPO to which the End User should direct requests in relation to their rights above;
    f. Retention period of the Personal Data Processed;
    g. Reference to this Policy.
8. Personal Data may only be transferred to or otherwise shared with individuals or organizations where the recipient:
    a. has agreed to be bound by this Policy and the set of common Infrastructure policies, or
    b. is part of a recognised Computer Incident Response Team framework and as part of an incident investigation to prevent active or suspected misuse of Infrastructure services, or
    c. presents an appropriately enforced legal request.

A.6 REFERENCES

R1	Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (DPD). http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046
R2	Approved EGI Security Policies. https://wiki.egi.eu/wiki/SPG:Documents

A.7 Infrastructure Participant Example Privacy Policy

This section provides an example of a privacy policy as required by the section A.5.7 above. It does not form part of the Policy on the Processing of Personal Data.

A.7.1 Privacy Policy

This Privacy Policy explains how we, *[insert Participant name here]* ("We"), treat data by which you can be personally identified ("Personal Data") as a result of your registration for and use of *[insert Infrastructure name here]* ("Infrastructure").

We collect the following Personal Data to identify you to enable us to grant you access to the Infrastructure and the services such as compute, storage and network that its participants offer:

- Name
- Email address
- Affiliation (e.g. VO)
- Certificate Distinguished Name (DN)
- *[Add or remove data as appropriate]*

To enable the Infrastructure to be safe and reliable for your use and to preserve your rights as a user we adhere to The Policy on the Processing of Personal Data ("The Policy") available here: *[insert url to PPPD here]*.

Your Personal Data will be shared but only where:

1. The recipient has agreed to abide by The Policy, or

2. Doing so is likely to assist in the investigation of suspected misuse of Infrastructure resources.

Your usage of the Infrastructure will be monitored. Records of this use, containing your Personal Data, may be shared as described above for operational, security and accounting purposes only. These records will be purged or anonymised after, at latest, 18 months.

You can contact our Data Protection Officer (*[insert contact details here]*) to obtain a copy of your Personal Data, request that it is corrected in case of factual error or if you suspect that it has been misused. You can also request that we stop using your Personal Data but this will affect your access to the Infrastructure.

This Policy should be read with reference to the Policy on the Processing of Personal Data and other Infrastructure policies available at *[insert link to Infrastructure Policies here]*.

[Insert Name and Contact Details of Infrastructure Participant]

Glossary

AA	Attribute Authority. An instance where attributes can be stored and retrieved based on a supplied identity. These attributes can be used with other information and an identity to make an authorization decision for that identity.
AAI	Authentication and Authorization Infrastructure.
DP CoCo	Code of Conduct. A formal agreement amongst partners.
GDPR	General Data Protection Regulation. A Regulation by which the European Commission intends to strengthen and unify data protection for individuals within the European Union.
IdP	**I**dentity Provider is a party that manages identities and provides an interface to those identities.
PII	Personal Identifiable Information. Information that can be used or combined with other data sources to lead to personal information like name, sex, religion, address, etc.
REFEDS	The Research and Education FEDerations group, which tries to articulate the mutual needs of research and education identity federations worldwide. https://refeds.org/
RI	Research Infrastructures.
SP	Service provider is a party offering one or more services to external users. For example, compute services.