Policies for e-Infrastructures

January 2016

“Policies for e-Infrastructures”

Contents

Introduction

The various organisations participating in an e-infrastructure are likely to have their own policies on its use; harmonising those policies offers an opportunity to implement them more accurately, efficiently and effectively.

This paper discusses how policies are likely to interact and how those developing policies can benefit from the coordination provided by using a common infrastructure.

Organisations may play a number of different roles in the e-infrastructure. The paper reviews which roles are likely to impose and implement policies, and the different areas that each policy is likely to cover (for example user authentication or physical security). Mapping the relationship between these policies should assist principal investigators, who appear responsible in practice for ensuring that data and infrastructures are used as intended. It should also help those developing and implementing policies to identify where else in the infrastructure similar concerns are likely to arise, providing opportunities both to choose common policy approaches that increase the likelihood of adoption and, in many cases, to have their policy requirements implemented and enforced more effectively by others.

E-infrastructure policies

Each of the components of an e-infrastructure is likely to have its own policies, whether these are referred to as acceptable use policies, terms and conditions of use or participant agreements.

Infrastructures and the communities that use them may have their own policies that providers and users are required to satisfy. In some cases these policies may cover the same issue: for example a data provider may have a policy that some datasets must be stored in secure facilities, infrastructure operators will have policies on operational security. An e-infrastructure may state that it is not to be used for personal data. The distributed nature of e-infrastructures – data from one organisation being processed by individuals from others using equipment provided and controlled by a third – means that compliance with these policies will often inevitably depend on the behaviour of others. Rather than representing a risk, however, this inter-dependence can create substantial benefits.

If, for example, the data provider's security requirement is matched by the infrastructure provider's provision then the data provider is saved the detailed checking of compliance with requirements that would otherwise be needed. Furthermore, the local infrastructure provider is likely to be much better placed to identify and remedy any policy failings than the remote data provider would be. Similarly if an infrastructure provider requires that their services are only used by current staff and students, users' home organisations are the best possible source of information of whether their current status satisfies that policy requirement. It would be both wasteful and inaccurate for the provider to attempt to replicate this information. Data and infrastructure providers who adopt common approaches increase the likelihood that users and their home organisations will have already committed to abide by and enforce those policies. Thus matching policies, accompanied by appropriate agreements and incentives, can both save the duplication involved in every organisation attempting to enforce the whole of its own policy and often result in more accurate and effective implementation by the authoritative source of information or control.

Conversely if a participant sets a unique policy requirement that is not addressed by any other organisation then they are likely to have to both implement the requirement and monitor compliance themselves. In a distributed infrastructure, where the policy-setting organisation does not have complete control or information, this may prove difficult or impossible. Ultimately the choice may be between accepting a policy that the infrastructure as a whole can implement accurately or keeping all data and processing in-house and trying to exercise complete control over access and use. The latter means giving up the significant benefits of e-infrastructure use so organisations following that course need to be sure both that their policy requirement justifies that loss and that the in-house solution will, indeed, implement it more accurately.

Some policy requirements will depend solely on the end user's conduct (for example whether commercial or unpublished use are permitted). Here there is unlikely to be anything any provider can do to **prevent** a policy breach, but policy matching may provide a more effective and dissuasive threat of sanctions if the policy is disobeyed. Provided individual agreements with end users are included in their home organisations' definitions of a policy breach, then a breach of policy could risk sanctions up to and including dismissal, rather than mere exclusion from future use of a particular infrastructure component.

As well as those directly involved in providing the e-infrastructure, other bodies may provide relevant policy contributions. These include, for example, the rules of access management federations, acceptable use and security policies of research and education networks, and professional codes of practice and accreditations. Bodies supervising these agreements may also provide an alternative way to resolve issues of compliance with their terms. Some research domains have developed Virtual Organisations (VOs) that individual researchers, principal investigators, data providers and infrastructure providers may join. VOs often require their members to agree to particular policies, reflecting agreed and accepted rules of behaviour in that domain. Where VO policies exist they provide a useful basis, and means of policy enforcement, for providers serving that domain.

It appears that principal investigators who are granted access to e-infrastructure data and resources are often, in practice, also delegated the responsibility for ensuring that their policies are complied with. Data providers are, however, likely to remain legally responsible, and all participants may suffer reputational consequences from a policy breach. However, outside the area of Certificate Policies, which may include some requirements on authentication strength and identity verification, there seem to be few automated tools available to help them. Matching policies would make the largely manual process of comparison significantly easier and less error-prone, increasing the likelihood that policy requirements will be delivered across all e-infrastructure components.

Who sets policies on what?

Organisations participating in an e-infrastructure can play one or more of three roles: as providers of data, systems or users. Each of those roles has an associated set of policies that the provider will require others to comply with as a condition of their participation.

Thus a data provider will normally wish to impose rules on how their data are used: depending on the nature of the data this may range from a requirement for open publication and preservation of results and methods, via whether or not commercial exploitation is permitted, to detailed requirements on who may access data and in what kinds of physical location. A provider of systems – whether compute, storage, connectivity or experimental equipment – is also likely to want to impose rules, including on the types of data and program that may be stored or processed on their equipment. A provider of users (in practice a provider of user authentication, normally the user's home organisation) is likely to impose rules on what information about those users may be processed or disclosed and for what purposes. Where an organisation performs multiple roles, such policies create requirements within that organisation, as well as between it and others. Although each of these policies may in practice be contained in a number of separate documents, for simplicity they are treated here as each comprising a single document, referred to as, respectively, the Data Policy (relating to the data provider role), the Infrastructure Policy (relating to the system provider role), and the Authentication Policy (relating to the user provider role).

Within an e-infrastructure these policies need no longer be isolated documents, since the policy and design choices of one provider may ensure (at least to an acceptable level of risk) that particular requirements of another provider are automatically satisfied. Thus, for example, a data provider's Data Policy requirement that information may only be accessed within a secure building can be satisfied by selecting service providers whose Infrastructure Policies state that they provide physical security to the appropriate standard. Considering each of these policies in turn, and examining the areas they are likely to cover, suggests many other situations where a requirement imposed in one role's policy is likely to be satisfied by the policy of another.

Data Policy (data provider role):

- May require:
  - Identity/status vetting: for example, that the individual is a member of a research organisation, or an approved researcher
  - Location: for example that the data will be held in a safe haven or a particular country
  - Physical/logical security: that data will be protected by particular physical and digital measures, or against a certain level of risk
  - Sanctions: that individuals can be held accountable for any breach of the policy
  - Freshness: that information provided about individuals is up-to-date
  - Conditions of use: for example that data may not be published, only used for non-commercial research, research published openly, results curated, etc
  - Incident response: that measures are in place to detect, contain and remediate any breach of security or policy

Infrastructure policy (service provider role):

- May deliver:
  - Physical/logical security
  - Location
  - Incident response

- May require:
  - Type of work: for example that the infrastructure may only be used for research in a particular domain, or for publicly-funded work, etc
  - Credential strength: that the means of authentication satisfy given technical, process and behavioural requirements (e.g. multi-factor, password length, non-sharing, expiry/revocation) or protect against particular forms of attack
  - Sanctions
  - Incident response

Authentication policy (user provider role):

- May deliver:
  - Credential strength
  - Identity/status vetting
  - Sanctions
  - Freshness
  - Incident response

- May require:
  - Physical/logical security
  - Incident response

Terms and conditions

Where one policy makes a requirement that is not satisfied by another, the most common approach is to include it in a set of terms and conditions that every individual user must agree to before being given access to the relevant e-infrastructure components. Terms and conditions may also reinforce requirements that are covered by another provider's policy: for example, a requirement that information be processed in a secure location may be reinforced by having the user personally commit not to remove it from that location. Terms and conditions may be imposed by any of the e-infrastructure roles: the analysis above suggests the following as the most likely areas for them to cover:

- Conditions of data use (e.g. that data will only be used for non-commercial purposes, that results will be managed and published in a particular way, confidentiality requirements)

- Conditions of service use (e.g. that a service will only be used for particular purposes, or in a particular domain of research)

For some high-sensitivity datasets terms and conditions directly agreed by the individual user may be the main part of the agreement. Whenever an e-infrastructure component requires its users to agree to terms and conditions, it should ensure that the user's home organisation, or some other party with a close relationship to the user, is willing to assist in investigating any breach of those terms and applying appropriate warnings or sanctions. The infrastructure component, being remote from the user, is unlikely to be able to apply effective sanctions itself. Even attempting to exclude a user from an infrastructure may be ineffective if the home organisation is not aware of the reason for this and simply "fixes" the user's problem by issuing them with new credentials.

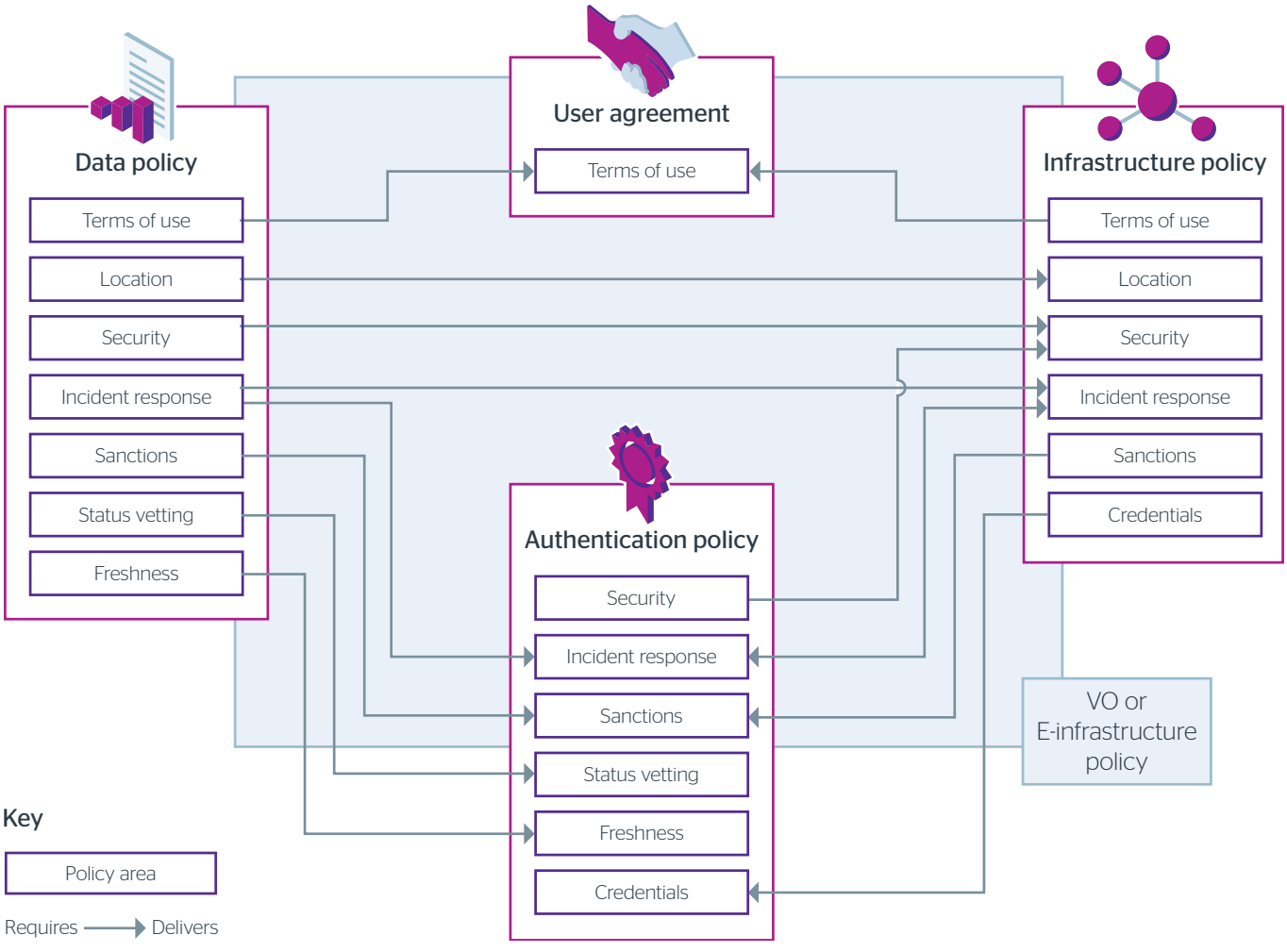
Virtual organisations

Virtual Organisations (VOs) can help coordinate both the setting and enforcement of policies. While some VOs simply gather together users, data and services in a common authorisation infrastructure, others specify in addition (with varying degrees of formality) the policies that their members should abide by. E-infrastructures may establish their own rules, covering any or all of these policy areas, for providers who wish to participate. In this way a Virtual Organisation or e-infrastructure that includes multiple projects and services may effectively establish the common set of policy requirements for work in a particular research domain. A VO or e-infrastructure may write its own policies or identify requirements (for example on security) that others'

policies should meet; it will often include external policies (for example the Janet AUP) as requirements for its members. In creating and using this "policy stack" the VO may identify inconsistencies or contradictions that policy setters should address.

Virtual Organisations may also provide ways to communicate policies and changes among their members. For example, some require all members to periodically re-confirm their acceptance of current policies; some inform members of policy changes or require them to check regularly that they are still complying with the current policy set; some may remind a user when they request an action (e.g. downloading a dataset) with significant policy requirements. VOs may also offer mechanisms for reporting and resolving any breaches of their policies, or at least communicating these among PIs and to users' home organisations.

The diagram below shows how these different policy areas are most likely to interact.



Policy enforcement

If one organisation relies on another to ensure that its policy is satisfied, it is reasonable to consider what will happen if that party fails to deliver what was expected of it. In the commercial world such breaches are most often dealt with by a monetary payment, but in a research infrastructure a payment may not be a useful way to resolve the problem. Unless the relationship has irretrievably broken down, the parties are likely to need to continue working together so any agreement should include a willingness to resolve problems by discussion. It will often be in both parties' direct interests to resolve the problem in any case – for example, a failure of an authentication process is likely to harm a home organisation's own systems at least as much of those of an infrastructure provider that relies on it. If difficulties arise then there will often be related parties, such as Virtual Organisations, research networks or federations (see the list of Common Policy Approaches, below), whose policies and processes can be used to reach a satisfactory outcome. Ultimately it may be technically possible for any party in an e-infrastructure to refuse to deal with any other but, since such action is likely to damage both parties, it should be considered only as a last resort. Temporary suspensions of access may be a useful incident response measure to contain the impact of a problem, but should always be accompanied by a clear process to notify relevant parties, resolve the root cause, and restore full normal operations.

Setting policies

Distributed e-infrastructures offer a rare opportunity: to have your policies enforced by someone who is better placed to do it, for free!

Conversely e-infrastructures may also work in ways that conceal, or even encourage, non-compliance with unrealistic policy requirements. When developing policies, those participating in e-infrastructures should obviously aim for the former outcome rather than the latter. This means setting policies that both users and other infrastructure components accept as reasonable to protect particular data or services, that will not encourage circumvention.

This should be helped by the fact that many risks apply in similar ways to different e-infrastructure components. For example, compromised accounts and users who do not respect terms and conditions are a threat to data providers, infrastructure providers and home organisations. This creates a common interest in taking measures to reduce the occurrence of these incidents and deal effectively with them when they do occur. Where a provider is already implementing a policy in its own interests this should give others confidence that it will be done well. When setting policies in a particular area, organisations should use the mapping in the previous section to identify both which other organisations may set policies in the same area and which may already have measures in place to satisfy them. Provided those organisations are concerned with the same risks as you are, adopting the same approach will normally offer the greatest benefits. Policies will be more likely to match if they state the aims to be achieved rather than the means that must be used: be prepared to accept others' similar policy goals and best practices, rather than insisting on identical wording. This may involve a careful study of the actual risks to ensure an appropriate approach – for example if the risk is that authenticated users will misbehave then strengthening the authentication process is unlikely to reduce it.

In some cases, however, a risk may appear specific to a particular dataset, service or home organisation. Where these are identified, organisations should first try to find others that are exposed to a similar risk, since there is a greater chance of other parts of the e-infrastructure making a change if it will satisfy multiple requirements. Also, where a common standard is required by a number of different services, those implementing it have a greater incentive to ensure compliance so as not to lose access to all those services. If your policy requirement is unique then it is harder to create a continuing shared interest with those who you must trust to implement it.

E-infrastructures serve a wide range of purposes, with very different policy requirements and expectations. For example, in theoretical research the only consideration is likely to be protecting individuals' priority of discovery; when investigating human subjects their privacy and other rights must be protected. Resources that are scarce or expensive to provide may require stronger policies to reduce waste. It is unlikely that all these requirements can be satisfied by a single set of policies: the access controls required to protect human subject data would, rightly, be considered intolerable by those modelling distant galaxies, and vice versa. However, the mapping approach suggested here should lead to the emergence of clusters of policies each addressing a particular sensitivity of research. For example, although user authentication can be done using a wide variety of processes and technologies, by considering what an attacker must do to successfully masquerade as a legitimate user (e.g. deceive a user, deceive a registration process, compromise an authentication system) these reduce to a much smaller number of distinct clusters each providing a different level of protection. Identifying and joining an appropriate policy cluster may take time,

given the number of slightly varying and overlapping policies that already exist, but is likely to provide far more effective protection and broader access than creating yet another unique variant. Within a cluster, matching policies will promote the use of distributed e-infrastructures and the greater efficiency and accuracy of policy enforcement that they can deliver.

Insisting on specialised requirements is likely to severely restrict the number of partners available: ultimately an organisation with such a requirement may need to implement it itself and accept that its data or service will only be available to the limited number of individuals who can comply and whose compliance can be verified. Requirements that are inappropriately hard to comply with risk creating incentives for users to adopt workarounds: there are many anecdotes of one member of a project team being assigned to obtain a 'high-strength' credential and then share it with the rest of the group. Such policy breaches may be particularly hard to detect when systems are designed to be used remotely and with devolved authorisation. Better protection may well be delivered by a less strict policy that users actually abide by than an apparently stricter one that is circumvented.

Since the principal aim of policy should be to reduce the occurrence of breaches, rather than to retrospectively punish those that occur, policy documents must be both clear and practical. Data and service providers must make clear to both principal investigators and users what behaviour will be expected of them and how that can be achieved. This must be done when contact is first made – whether by a request to use data or a service, or to join a project. Principal investigators should be told what policy areas they are responsible for and which existing policy approaches – either within a particular research domain or from external sources such as research networks or federations – are likely to be suitable. Using a policy structure that can be mapped to other parts of the e-infrastructure will help PIs identify appropriate tools and behaviours. Referencing policies that are already familiar should make it easier for users to comply, as well as highlighting any specific differences that require care when using a particular data set or service.

Common policy approaches

A number of policy approaches have already gained acceptance within research communities. Infrastructure components that adopt these as the basis for their own policies are likely to find most partners already familiar with them, and many already able to comply.

The areas likely to be covered by each of these are as follows:

- **UK access management federation** (provides a basis for authentication/authorisation to online services and resources): covers credential strength, identity/status vetting, sanctions, freshness. (see ukfederation.org.uk/content/Documents/FedDocs)
- **UK Government GPG45** (Identity proofing and verification of an individual): covers identity vetting (see gov.uk/government/publications/identity-proofing-and-verification-of-an-individual and the Working Group's paper on Authentication)
- **Grid authentication profiles** (issuing X.509 PKI credentials to end users): covers identity vetting and credential security (see eugridpma.org/guidelines)
- **CSC Top20 controls** (how to protect infrastructure components against attack across networks): covers logical security (see e-Infrastructure Security paper)
- **ISO/IEC 27001** (requirements for information security management systems): covers physical/logical security, credential strength, sanctions (see **UCISA information security management toolkit** ucisa.ac.uk/ismt)
- **NHS Information Governance Toolkit** (requirements for handling medical data): covers physical/logical security, credential strength, sanctions, location, identity/status vetting
- **Domain policies** (requirements for membership of a Virtual organisation or e-infrastructure): may cover any aspect of policy. See for example EMBL-EBI, EGI (egi.eu/about/policy/policies_procedures.html)

Whether your role is as data provider, system provider, user provider/home organisation or virtual organisation, choosing an appropriate set of policies – reusing existing policies wherever possible and only creating new ones where this is unavoidable – can greatly increase the benefits you obtain from distributed e-infrastructures.

Annex: Membership of the UK e-infrastructure security and access management working group

- **Stephen Booth**, Edinburgh Parallel Computing Centre
- **Peter Boyle**, Edinburgh University
- **David Britton**, Glasgow University
- **John Chapman**, Jisc
- **Andrew Cormack**, Jisc
- **Darren Hankinson**, Manchester University
- **Josh Howlett**, Jisc
- **Henry Hughes**, Jisc
- **Jens Jensen**, STFC
- **David Kelsey**, STFC
- **Paul Kennedy**, Nottingham University
- **Bridget Kenyon**, UCL
- **Philip Kershaw**, STFC
- **Steven Newhouse**, EBI
- **Tommi Nyronen**, CSC
- **Jeremy Olsen**, Francis Crick Institute
- **Alan Real**, Leeds University
- **Andrew Richards**, Oxford University
- **David Salmon**, Jisc
- **Jeremy Sharp**, Jisc
- **Rhys Smith**, Jisc
- **Melanie Wright**, Essex University
- **Ioannis Xenarios**, Swiss Institute of Bioinformatics

Share our vision to make the UK the most digitally advanced education and research nation in the world

jisc.ac.uk

Jisc
One Castlepark
Tower Hill
Bristol, BS2 0JA
0203 697 5800
info@jisc.ac.uk