

When a PET is a Chameleon

Author:

Andrew Cormack

JANET(UK), Lumen House, Library Avenue, Harwell Science and Innovation Campus,
Didcot, OX11 0SG

Abstract:

Federated Access Management (FAM) is a combination of technology and agreements that allows on-line services to provide access to authorised users without knowing their real world identities. When compared with traditional individual user accounts FAM can provide both more accurate authorisation decisions – for example where an on-line service is licensed to all students at a university – and greatly improved privacy protection. As such it has been recognised by privacy regulators as a Privacy Enhancing Technology. However legislation, advice and case law are unclear and sometimes contradictory on the legal status of similar technologies, creating a barrier to the use of Federated Access Management within, and particularly between, countries. The problem arises because current EU laws encourage regulators and courts to concentrate on a simple binary decision – whether processing involves personal data or not – rather than adopting a risk-based approach to protecting privacy. The binary approach provides no incentive to adopt privacy protecting techniques and is even leading courts to exempt privacy-invasive processing from regulatory oversight.

Federated Access Management

Federated Access Management (FAM) is a relatively new technology that allows users to authenticate themselves to an organisation and then have that organisation (acting as Identity Provider) make trusted statements about them to a service provider. The service provider can then use the statements to decide whether or not the user is authorised to use the service. For example when a law student wishes to access an on-line journal, they can log in to their university, the university can then inform the journal that the user is indeed a student, and the journal can grant or deny access depending on whether the university has a site licence permitting access to all students.

Comparing this with the more traditional approach where each user has an individual login account with the journal website highlights the benefits for both the user and the website. The user gets improved privacy because there is no need to reveal their identity to the journal. The journal gets accurate and up to date information about users' status so it can enforce its licence terms without having to maintain a large set of login accounts and trying to disable them as users cease to be eligible.

These benefits are well known. The UK Information Commissioner has recognised Federated Access Management as a Privacy Enhancing Technology;¹ at the time of writing the UK Access Management Federation for Education and Research connects 445 Service Providers and 635 Identity Providers; while the eduroam™² service can provide university staff and students with authenticated network access when visiting universities across Europe and the wider world.

Pseudonymous Identifiers

While many services can make authorisation decisions based only on characteristics of a user – for example that they are a recognised user or have a particular type of relationship with their organisation – it is often important for users to be able to store and recover information between different sessions on the same service. For example I might wish a search tool to remember queries I have entered on previous visits, or have a portal remember my preferred colour scheme or layout of screen tools.

For these types of application, most Federated Access Management systems can provide unique opaque identifiers that allow services to distinguish individual users and recognise them when they return, but not to link the user to any real world identity. Indeed many systems will provide a different identifier to each service to prevent services colluding to track a user or to aggregate information about them. Identifier values are normally assigned either randomly or using cryptographic hash functions to ensure that they do not contain any information that might reveal

¹

http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/privacy_enhancing_technologies_v2.pdf

² <http://www.eduroam.org/>

the individual's identity. Federation policies normally prohibit any service from attempting to circumvent this anonymisation. Since these identifiers are designed to be anonymous for the recipient Service Providers, but not for the Identity Provider that issues them, they are referred to as **pseudonymous identifiers**.

Although there do not appear to have been any court cases dealing specifically with pseudonymous identifiers, a number of cases have considered the legal status of Internet Protocol (IP) addresses. These have some of the same characteristics: in particular the identity of the person using a particular IP address at a given time will often (though not always) be known to the originating Internet Access Provider but not to the operator of a service that the user accesses. However IP addresses are less privacy-protecting than pseudonymous identifiers as they will often contain some information about the user's location and connectivity. Different services will normally see the same IP address for the user raising the possibility of combining information from multiple services to gain a larger, aggregated, set of information about the user. Cases on IP addresses can therefore provide some information about how the law is likely to treat pseudonymous identifiers, but are not a perfect model.

Personal Data?

Since the pseudonymous identifiers used in FAM systems may be associated with distinct individuals, this raises the question of whether they are personal data, subject to the UK's *Data Protection Act 1998*³ and the European Directive 95/46/EC.⁴ In the hands of the Identity Provider who issues them it seems clear that they are, since the Identity Provider will know which account logged in and which individual is the owner of that account. Indeed most Federation policies require the Identity Provider to be able to deal effectively with an individual responsible for any misuse of a resource or the access management system. However in the hands of a Service Provider who does not know either the login account or its ownership the position is much less clear.

The UK *Data Protection Act 1998*, s.1(1) defines

“personal data” means data which relate to a living individual who can be identified—

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller.

Whereas the European *Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*, 95/46/EC, Article 2(a) defines

“personal data” shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;

³ http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1

⁴ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML>

And in recital 27:

... to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person...

There has been some analysis of the word “relate”, following the UK case of *Durant v Financial Services Authority*.⁵ However for Federated Access Management, and pseudonymous identifiers in general, the more problematic words are in fact “can be”, “identified” and “likely”. Each of these will be discussed in turn.

“can be”, by whom?

Both UK and European law recognise two types of personal data: information such as name and address that has a *direct* link to a living individual (part (a) of the UK definition), and information that can only be linked by combining it with further knowledge (part (b) of the UK definition and “*Indirect identification*” in the Directive). Both laws, reasonably enough, seem to regard the first type as being fixed permanently as personal data no matter who holds it, but there appears to be a difference in the treatment of indirectly linked information. Is such information still personal data when held by a person or organisation that does not have the necessary further knowledge? The Directive suggests that it is, by taking into account actions likely to be taken “by the controller *or by any other person*”: this seems to imply that so long as it is possible for *some* person to make the link, then indirectly linked information, too, is fixed permanently as personal data whether or not the current data controller can make the link. By contrast UK law seems to allow the original information not to be personal data so long as it is held by someone who “is [not] likely to” gain possession of the linking information.

This latter interpretation has recently been explicitly confirmed in the Irish case of *EMI Records & Others v Eircom Ltd*.⁶ EMI had appointed an agent to monitor traffic on Eircom’s backbone network, scanning for illicit transfers of copyright music. When such a transfer was detected, the agent recorded the source IP address and identified the music concerned and passed this information to Eircom as a copyright breach report. The Irish Information Commissioner asked the High Court:

16... “Do data comprising IP addresses, in the hands of EMI or its agent(s), and taking account of the purpose for which they are collected and their intended provision to Eircom, constitute “personal data” for the purposes of the Data Protection Acts, 1988-2003, ...?”

The Irish *Data Protection Act 1988*⁷ has a very similar two-part definition to the UK Act, also in s.1(1):

⁵ *Durant v Financial Services Authority* [2003] EWCA Civ 1746 <http://www.hmcourts-service.gov.uk/judgmentsfiles/j2136/durant-v-fsa.htm>

⁶ *EMI Records & Others v Eircom Ltd* [2010] IEHC 108, <http://www.courts.ie/Judgments.nsf/09859e7a3f34669680256ef3004a27de/7e52f4a2660d8840802577070035082f?OpenDocument>

⁷ <http://www.irishstatutebook.ie/1988/en/act/pub/0025/>

"personal data" means data relating to a living individual who can be identified either from the data or from the data in conjunction with other information in the possession of the data controller;

The Information Commissioner and the Court assumed, reasonably enough, that this meant that the IP addresses were personal data when received by Eircom. As discussed above, it also seems obvious that the addresses must have been personal data when Eircom originally assigned them to the user. However when, in the interim, the addresses were held by EMI and its agents the judge concluded at paragraph 25 that "the ... question is therefore answered no" because "I do not regard it as at all likely that [EMI] will attempt in any way to use the IP address ... in order to find out their names and addresses". Indeed an earlier case (*EMI Records v Eircom Limited*⁸) had determined that the names and addresses could only be found using the information held by Eircom and therefore granted a *Norwich Pharmacal* order requiring Eircom to disclose that information. So in Irish (and presumably UK) law, the same indirectly identified information can change state, from personal data to non-personal data and back again, depending on who holds it.

However it is not clear that the same view is taken by other European countries. Six member states were studied by Hunton & Williams for a report on Online Copyright Enforcement for the European Commission.⁹ This asked a single question "Is an IP address personal data?", thus apparently ruling out the possibility that the answer might depend on who holds the IP address. However some countries' responses indicate that different considerations do apply to IP addresses when held by ISPs and rightsholders: for example Austria "Monitoring, investigation, filtering and processing by [rightsholders] of temporary IP addresses without additional information are not covered by data protection law since they do not result in the processing of personal data" (the same argument to the Irish court in *EMI v Eircom*). Other countries, including Belgium, consider the same copyright enforcement activity not only to involve processing of personal data, but to relate to criminal activity and therefore prohibited for anyone other than the judicial authorities! For Germany, the survey reports that the Data Protection Authority considers that "regardless of the type of data controller, IP addresses always constitute personal data". German courts have, however, been less clear when asked to rule on the status of web server logs that are indexed by IP address: in case 23 S 3/07 the Berlin Regional Court stated that these were personal data, whereas in case 133 C 5677/08 the district court of Munich stated that they were not.¹⁰

Another possibility is that whether an indirect identifier is personal data may depend on the current holder's intentions. EMI were considered to have no wish to make the identification themselves, so the Irish court concluded that they were not handling personal data. However this appears to conflict with the UK Information Commissioner's view that any information that "informs or

⁸ *EMI Records v Eircom Limited* [2005] 4 I.R. 148

⁹ http://ec.europa.eu/internal_market/iprenforcement/docs/study-online-enforcement_en.pdf

¹⁰ http://www.2b-advice.com/no_cache/service/meldungen/2b/news/2008/11/24/d-court-declares-ip-addresses-are-non-personal-data.html

influences actions or decisions which affect an individual”¹¹ constitutes personal data. Since EMI’s processing could lead to the individual losing their Internet connection, this UK approach would seem to lead to the opposite conclusion.

The answer to whether an indirect identifier is always personal data, or only in the hands of someone who can (or, perhaps, who wishes to) link it to the living individual, is therefore far from clear.

“identified” or recognised?

The copyright cases above seem to presume that for someone to be “identified” requires that their on-line activity be linked to their real world identity. For example *EMI v Eircom* refers at paragraph 25 to “names and addresses”, the Article 29 Working Party to “flesh and bone”.¹² However other opinions suggest that finding the real world identity is not necessary. Elsewhere in the same Opinion, the Article 29 Working Party state that “the possibility of identifying an individual no longer necessarily means the ability to find out his or her name”.¹³ Their view, both in the context of both “web traffic surveillance tools”¹⁴ and search engines¹⁵ seems to be that the ability merely to *recognise* a series of events as being performed by the same person, even if there is no possibility of finding out who that person is, is sufficient.

If this is correct, of course, then the question of all pseudonymous identifiers could be answered very simply since their whole purpose is to permit recognition between different visits to the same service, while doing everything possible to prevent identification of the individual involved. However, as noted above, an IP address can be regarded as a rather poorly protected pseudonymous identifier – an IP address will often (though not always) allow the recognition of a returning visitor. So if recognition is sufficient, then the Irish and other copyright cases could have concluded much more quickly and with the opposite result.

In fact the Article 29 Working Party themselves give a counter-example where recognition is not sufficient to create personal data.¹⁶ They describe a medical trial that allocates each patient an opaque identifier protected against re-linking by both technical means and a legal agreement, and conclude that the trial data labelled with this identifier can be treated as non-personal data.

¹¹ Data Protection Technical Guidance: Determining what is personal data, p.9, http://www.ico.gov.uk/upload/documents/library/data_protection/detailed_specialist_guides/personal_data_flowchart_v1_with_preface001.pdf

¹² Opinion 4/2007, p.13, http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf

¹³ Opinion 4/2007, p.14

¹⁴ Opinion 4/2007, p.14

¹⁵ Opinion 1/2008, http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_en.pdf

¹⁶ Opinion 4/2007, p.20

It therefore seems that depending on other factors, which are not entirely clear, sometimes “recognition” is sufficient but in others “identification” will be required.

How “likely”?

Both UK and EU laws contain a test of whether it is “likely” that information will be linked to an individual, but there appears to be a wide range of opinions on how strong the likelihood has to be and what factors should be taken into account.

The Irish court in *Eircom* seems to use the narrowest definition of personal data, stating at paragraph 23 that “likely means probable”, and assessing this probability based only on the past conduct and presumed future intention of the Data Controller. In fact a deep-packet inspection system capable of identifying copyright files is examining network traffic at exactly the same level that could also reveal e-mail addresses, web or social network postings containing direct personal identifiers associated with the same IP addresses. The system is presumably not programmed to extract that linking information, but there seems no technical reason to prevent it doing so.

The two German logfile cases take wider definitions. Both courts agreed that it was possible for the owner of the website to obtain the linking information. Munich determined that it was not “likely” that they would do so because this would require a breach of privacy law by the relevant Internet Access Provider, whereas Berlin concluded that the need for an unlawful act did not take the possibility outside the definition of “likely”.¹⁷ Interestingly, neither court considered whether the logfiles themselves were “likely” to contain information, for example a unique pattern of accesses, that was itself sufficient to identify the individual user, though this does seem to be the Article 29 Working Party’s principal concern with the logs of search engines.¹⁸

The Article 29 Working Party take the strictest view, that identifiers will be personal data unless the possibility of linking “does not exist or is negligible”:

If, taking into account “*all the means likely reasonably to be used by the controller or any other person*”, that possibility does not exist or is negligible, the person should not be considered as “identifiable”, and the information would not be considered as “personal data”.¹⁹

The Working Party do concede that “likely” requires something more than a “mere hypothetical possibility”,²⁰ but, when considering the example of website logs indexed by IP address, state that these should be treated as personal data unless the website operator knows “with absolute

¹⁷ http://www.2b-advice.com/no_cache/service/meldungen/2b/news/2008/11/24/d-court-declares-ip-addresses-are-non-personal-data.html

¹⁸ Opinion 1/2008, p.4

¹⁹ Opinion 4/2007, p.15

²⁰ Opinion 4/2007, p.15

certainty” that it is impossible for them to be linked, for example because they are associated with a wireless network with no registration requirement.²¹

Depending on the source, therefore, it appears that “likely” can mean anything from “not impossible” to “probable”.

What About Pseudonymous Identifiers?

From the discussion above it seems that, sometimes, an IP address is not personal data. Since IP addresses are not particularly designed to protect privacy it seems that the same should apply to the pseudonymous identifiers used by many Federated Access Management systems, which do use both technical and procedural measures to reduce the likelihood of identification. Indeed the same technology – one-way cryptographic hashes – is used in the drug trial example given by the Article 29 Working Party as an instance of non-personal data.²²

Two of the options above must therefore be incorrect: either that recognition is always sufficient – because IP addresses and pseudonymous identifiers almost always provide that – or that it is sufficient that *someone* can make the link to identify the individual – because the Internet Access Provider or Identity Provider who issues the identifier can almost always do that. Either of those options would mean that both IP addresses and pseudonymous identifiers were *always* personal data.

Since both IP addresses and pseudonymous identifiers are personal data at the time when they are assigned to a known person, this leads inevitably to the conclusion that they must be able to change state, becoming non-personal data once they have travelled sufficiently “far” from their origin for identification to be unlikely. The reverse change, from an IP address being non-personal data to personal data, is explicitly recognised in *Eircom*, which implies an earlier change from personal to non-personal somewhere on the address’s journey from Eircom to the user to the monitoring device.

This seems an attractive solution: that by disclosing an identifier without its linking information the Identity Provider can render it non-personal. Unfortunately two aspects of this change of state are not addressed by current law: what happens if the state change takes place during transfer from within the EEA to outside, and what happens if the recipient of a non-personal pseudonymous identifier subsequently obtains sufficient information to link it to an individual, either from the individual or a third party or by aggregating information associated with the same identifier?

International transfers

European Data Protection law places particularly strict requirements on any transfer of personal data from inside the European Economic Area to outside. Article 25 of the Directive requires a data controller who discloses information to ensure that it continues to benefit from European-standard protection, even if the country to which it is transferred does not provide that protection in its own

²¹ Opinion 4/2007, Example 15 on p.17

²² Opinion 4/2007, p.20

legislation. However the law does not seem to envisage the possibility that the information may change state during this transfer. If information is personal data in the hands of a European Identity Provider, but not in the hands of the non-European Service Provider to whom it is transferred, then does Article 25 apply or not?

This might seem an obscure question, but following *EMI v Eircom*, it is one that arises billions of times a day, whenever a European Internet Access Provider sends a packet, labelled with an IP source address, to a server outside Europe! By that judgment the source address is personal data in the hands of the ISP, but quite possibly not in the hands of the foreign website that receives it. If the Internet Access Provider is required to abide by Article 25 for this transfer of personal data, it is not clear how it can lawfully send the packet at all, since it will not normally have any agreement with the organisation operating the recipient server! It might be argued that the user consents to the release of this information, but how can that consent be “informed” when techniques such as Anycast IP addresses²³ mean that the destination of the packet, in particular whether it is inside or outside Europe, may not be determined until the instant the user clicks their mouse?

Additional Information

A further problem arises if a pseudonymous identifier is released to a Service Provider under circumstances that would normally make it non-personal data, but the Service Provider then obtains information from some other source that allows them to link it to a particular individual. This may be deliberate, most obviously if the Service Provider invites the user to provide personalising information directly, or accidental if the user discloses their identity (doing so deliberately has been suggested as a way of ensuring that on-line activity is protected by the *Data Protection Act*²⁴) or if their activity on the service creates a unique identifying pattern. If this happens, is the Identity Provider retrospectively responsible for having released something they had in good faith considered to be non-personal information, or are they required to stop future releases until full regulatory compliance has been restored?

Again, the circumstances of *EMI v Eircom* offer interesting possibilities. What if the monitoring agent’s equipment were found to be collecting packet contents that happened to include e-mail addresses or other identifiers (as has recently been discovered for Google’s Streetview service^{25,26})? Or what if the intended monitoring of copyright downloads found a unique pattern of musical interests that could be associated with an identifiable single individual? In fact this appears to be the Article 29 Working Party’s concern: not that IP addresses are themselves privacy invasive, but that using them as a label might create the ability to “categorise [a] person on the basis of socio-economic, psychological, philosophical or other criteria and attribute certain decisions to him”.²⁷

²³ <http://en.wikipedia.org/wiki/Anycast>

²⁴ Pounder, C “Reclaiming Privacy on the Internet”, July 2009, <http://www.amberhawk.com/uploads/IPSTREETVIEW.pdf>

²⁵ <http://www.identityblog.com/?p=1120>

²⁶ <http://news.bbc.co.uk/1/hi/technology/10364073.stm>

Cases that exempt copyright infringement monitoring from personal data regulation seem to be increasing precisely that risk.

Potential Personal Data?

Law, cases and opinions all seem to indicate that pseudonymous identifiers may both be and not be personal data. The answer seems to depend on the question's background, like a chameleon! Relevant factors seem to include the identity of the person holding the information, their presumed intentions and even the courts' view of how technologies may develop in future.

This shifty status, combined with the significant regulatory change between "personal" and "non-personal" states, is not helpful either to privacy or to those designing and operating systems using pseudonymous identifiers. On the one hand it is clear that some uses of these identifiers can be as great a threat to privacy as direct identifiers so should be subject to the same regulation; however other uses can be significant privacy improvements (as recognised by the classification of Federated Access Management as a Privacy Enhancing Technology) and ought to be encouraged. Unfortunately the lack of clarity as to what regulation applies may actively discourage their use: system designers may well prefer to use direct identifiers such as names or e-mail addresses whose regulatory status is at least clear.

It may therefore be that "is this identifier personal data?" is actually the wrong question to ask. Instead any identifier that might permit recognition could, perhaps, be considered Potential Personal Data: regulated but with the strictness of regulatory requirements depending on how great a risk the identifier's design and processing presents. Thus using an IP address to build up a pattern of use across multiple services might be subject to the same data protection requirements as a direct identifier, whereas much lighter regulation might be applied to an opaque identifier that is protected against aggregation and de-anonymisation by both technical and contractual means and where the user controls what information is stored against it.

This may be what the Article 29 Working Party had in mind in introducing their Opinion 4/2007 on the Concept of Personal Data:

The Directive contains a broad notion of personal data...²⁸

...It is a better option not to unduly restrict the interpretation of the definition of personal data but rather to note that there is considerable flexibility in the application of the rules to the data.²⁹

It might also provide a consistent reason for their apparently contradictory views that, in an example of a drug trial:

²⁷ Opinion 4/2007, p.14

²⁸ Opinion 4/2007, p.4

²⁹ Opinion 4/2007, p.5

The data do not contain any additional information which make identification of the patients possible by combining it. In addition, all other measures have been taken to prevent the data subjects from being identified or becoming identifiable, be it legal, technical or organizational. [And therefore they are **not** personal data]³⁰

While at the same time for web server logs containing dynamically allocated IP addresses:

The same can be said about Internet Service Providers that keep a logbook [sic] on the HTTP server. In these cases there is no doubt about the fact that one **can** talk about personal data in the sense of Article 2(a) of the Directive.³¹

Using a concept such as Potential Personal Data, *EMI v Eircom* and other copyright cases would ask not whether or not the processing involved personal data but instead whether that processing of potential personal data was justified by a legitimate interest of a third party, and whether or not this interest was over-riden by the fundamental rights of the data subject.³² Such analysis would create an incentive to design and use privacy-protecting systems that would be more likely to pass the second part of the test, rather than the current incentive to try to exclude regulation entirely.

Such an approach should also recognise that separating identifiers from their linking data can protect privacy by separating knowledge of what was done from knowledge of who did it. Organisations that implement such separation of duties should be rewarded rather than, as at present, having to justify both halves of the processing **and** the transfer between them.

Unfortunately this flexible approach is not what current law, with its binary choice between regulated personal data and unregulated non-personal data, provides. Some of the Directive's requirements for processing and handling personal data do have "reasonableness" tests, for example Article 11(2) waives the requirement to notify the user if "the provision of such information proves impossible or would involve a disproportionate effort", and Article 17(1) requires "appropriate technical and organizational measures" to protect the security of the personal data (though regulators' advice rarely seems to take into account how well different types of "personal data" may already be protected). However the requirements on international transfers (Article 25) and Subject Access Requests (Article 12) seem to apply to everything classed as personal data, even though verifying the identity of the person making the SAR may well be impossible if only an indirect identifier, such as an IP address or pseudonymous identifier, has been used to index the "personal data" whose disclosure is required! This may be one reason for the UK Information Commissioner's admission that the current law poses "practical difficulties, sometimes insurmountable ones, in complying with all aspects of the DPA in respect of non-obvious personal identifiers!"³³

³⁰ Opinion 4/2007, pp15-16

³¹ Opinion 4/2007, p.16

³² Recital 30 of Directive 95/46/EC

³³ Personal Information Online Code of Practice, Consultation Document (Dec.2009) p.4, http://www.ico.gov.uk/upload/documents/library/data_protection/practical_application/pio_consultation_200912.pdf

Conclusion: Being Nice to PETs

The fact that both courts and regulators are having such difficulty answering the apparently simple question “is an IP address personal data?” suggests that it may in fact be the wrong question. Although IP addresses did exist in 1995 when the European Directive was passed, it seems highly unlikely that the drafters were sufficiently aware of them to foresee the problems they cause for the definition and regulation of indirectly identifying personal data. The problems are significant, because the current state of confusion appears both to be permitting some privacy-invasive activities to take place outside the scope of regulation and to be discouraging (because of the legal uncertainty) the adoption of technologies that can significantly improve privacy. Some interpretations of current law appear to recognise no difference between an e-mail address that directly identifies the user and a unique, opaque pseudonymous identifier that cannot be linked to a user without a breach of contract and where the user controls what additional information is associated with the identifier.

The problem for courts is that current law and practice seem to contain only a crude binary distinction between personal data and non-personal data. Once information has been placed in one or other category some of the legal requirements are absolute – in particular those on Subject Access Requests and international transfers – and even where the law appears to allow flexibility by using terms such as “appropriate” and “reasonable”, advice from regulators too often replaces this flexibility by absolute requirements. When classification as personal data can create “insurmountable difficulties” to performing an activity in compliance with the law, the courts are forced to pick the least bad of the two options for each particular situation.

Resolving this problem seems to require changes both to the law and to regulators’ approach to it. Indirectly linked information needs to be recognised as a separate category – referred to above as “potential personal data”. Depending on the circumstances some types of processing of this class can be as intrusive as directly-linked data while others can be effectively anonymous; law, regulation and advice need to recognise this range of possibilities with flexible requirements that promote the use of less privacy-invasive techniques and technologies. Such an approach in fact follows the Article 29 Working Party’s invitation to “the development of a policy that combines a wide interpretation of the notion of personal data and an appropriate balance in the application of the Directive’s rules.”³⁴ Determining and documenting that appropriate, risk-based, balance will be more complex than declaring blanket rules for all personal data – transferring information to a third party may be privacy-invasive if it facilitates aggregation or privacy-protecting if it separates knowledge of what was done from who did it – but it needs to be done. A complex system that recognises the privacy opportunities and risks of current and future technologies is preferable to a simplistic system that produces such incoherent results that it has to be ignored.

³⁴ Opinion 4/2007, p.5