Brian Pickering, Callum Beamish, Clare Hooper, Mike Surridge
{jbp|cab|cjh|ms@it-innovation.soton.ac.uk}

Multiple data owners: *who's doing what with your data?*

The Future Internet offers increasing opportunities for participation by private individuals, *natural persons* in legal terms[1]. Personal access devices have not been confined to office-based personal computers for some time, and continue to evolve: computer systems grew smaller and more compact with a demand for increased portability, and personal communication devices (mobile phones) grew in storage and processing capacity as well as going beyond telecommunications to the web (smart phones) for the two to converge in tablet-type devices. On the one hand, this allows for extensive and pervasive connectivity all day, every day, for access to data and information systems, to communicate with friends, with colleagues and with businesses and government, as well as to share with the world or worlds what us going for the individual or in an individual's reaction to events or to others: the social network. On the other, this poses increasing challenges for personal privacy as well as freedom. Personal data associated with individuals should be treated with care, it can be assumed; but what happens when the data subjects themselves release such data via social networking sites (SNS)?

In this report, relevant legislation surrounding the treatment of personal data is presented and reviewed. Interactions of individuals (*data subjects*) with online services is described against the legislative background and summary conclusions and recommendations are made directed at FI *Users*, FI *Providers* and *Service and application developers.* The report is divided into the following sections:

> *Background: the legal perspective on protecting personal data* outlines the legal framework in Europe for the protection of personal data, summarising the various sections of the Data Protection Directive for how such data should be handled.
>
> *The reality: should we be nervous?* discusses how legislation is implemented and lists areas such as unauthorised disclosure and sharing in terms of particular cases against well-known service providers.
>
> *User perceptions: trust* briefly reviews user attitudes to online services and how their personal data are protected.
>
> *User confidence: the public domain* outlines the legal basis for treating data which have been made public (such as varying sharing on public websites); and finally
>
> *User profiles and data mining: derivative works* looks how personal data shared via social networking sites along with records of online activity and behaviours can be used to build up profiles of end users which could well provide an unwanted perspective on a given individual.

So the intention in this overview is to bring together legislative, subjective and service-oriented aspects of personal data usage as it stands today with some indicators of the challenges for those building as well as using the Future Internet.

[1] A *living* person as is commonly understood; the yet unborn and dead are not afforded the same legal status.

Background: *the legal perspective[2] on protecting personal data*

Issue	The FI offers unprecedented access to data sources, with smart devices and sensors able to collect information dynamically to be transferred on for aggregation and interrogation as well as services and applications dedicated to interactions between users and whole communities both locally and across whole regions and continents. In such an environment, the issue of personal data – not just personal identifiers such as name, address and so forth, but also location, online activity and behaviours – becomes increasingly relevant. How the data should be dealt with must first consider existing legislation and its implications for service developers, infrastructure providers and of course end users themselves.

The Data Protection Directive (DPD)[3] establishes a framework which aims to protect the "fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the *processing of personal data*"[our italics]. Under Article 2(a) the concept of "*personal data*" is defined as

> "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".

The processing of such personal data is a major focus of the DPD and is described at Article 2(b) as

> "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction".

This is particularly relevant to all FI ecosystem stakeholders, and especially to the service and application developers and providers, as well as the infrastructure owners who provide the resource to run those services and applications.

With a relational framework established between *data subject* (as defined above), *data controller* (the person or organisation deciding how and why data should be processed) and the data processor (the person or organisation actually carrying out any such processing), the DPD stipulates particular principles that must be adhered to when data are being processed, as found under Article 6, and require that personal data must be:

(i) "processed fairly and lawfully;

(ii) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes […];

[2] Throughout this document, and for the purposes of simplicity, we have concentrated principally on EU law, or that of its member states.

[3] Council Directive 95/46/EC 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 (1995) OJ L 281/31

(iii) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;

(iv) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;

(v) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed […]".

And it is on those terms that *consent* should usually be sought: the data subject needs to know what the data will be used for, and have a right to access the data to check validity and currency.

At the EU level, the data protection framework provided for in the DPD does not distinguish between personal data that is in the public domain until that data falls under Article 8 as "special categories of data".

Article 8(1) establishes that no processing of personal data which reveals "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership" shall occur and nor shall "processing of data concerning health or sex life". This statement is then subject to certain exclusions listed in Article 8(2) as

> "(a) the data subject has given his explicit consent to the processing of those data […]; or
>
> (b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or
>
> (c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
>
> (d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or
>
> (e) the processing relates to data which are *manifestly made public by the data subject* or is necessary for the establishment, exercise or defence of legal claims." [our italics]

This is the only part of the Directive which distinguishes between data not in the public domain and data which is. There is no explicit EU law on what "manifestly made public by the data subject" means, but it is generally understood as requiring "a deliberate act by the data subject, disclosing the data to the public". For example, therefore, video surveillance would not be considered appropriate, but an interview to media, or publication on a public internet page would make data public.[4]

[4] Kotschy in Alfred Büllesbach et. al., *Concise European IT Law,* (2010, 2nd Edn, Kluwer Law International) 62

Following the absence of discussion regarding public domain throughout the rest of the Directive, it is clear that where personal data exists in the public domain and is subsequently processed, this processing is still subject to the Directive's principles and requirements.

Indeed, if one looks to the UK Information Commissioner's Office (ICO), they noted in their online code of practice[5] that:

- "People may post their personal details in such a way that they become publicly visible – for example through a social networking or recruitment site. *Wherever the personal data originates, you still have an overarching duty to handle it fairly and to comply with the rules of data protection*[…]

- If you collect information from the internet and use it in a way that's *unfair or breaches the other data protection principles*, you could still be subject to enforcement action under the DPA even though the information was obtained from a publicly available source.

- It is good practice to only use publicly available information in a way that is unlikely to cause embarrassment, distress or anxiety to the individual concerned. You should only use their information in a way they are likely to expect and to be comfortable with. *If in doubt about this, and you are unable to ask permission*, you should not collect their information in the first place." [our italics]

Thus, if personal data is being processed that was collected from the public domain it must adhere to the data protection principles[6] (see (i) to (v) above). In essence, therefore, the final point of the ICO guidance is most important. The use must be "unlikely to cause embarrassment, distress or anxiety to the individual concerned" and a use of "their information in a way they are likely to expect and to be comfortable with". This guidance implies an element of consent is granted through a data subject making that data publicly available, making the processing "lawful", and the controller must then ensure that the use does not cause embarrassment, distress or anxiety which meets the "fairness" requirement.

So nominally, FI users/participants should be adequately protected from embarrassment and the misuse of their data, not least since the *natural* or *legal person*[7] responsible for controlling and processing any data collected by a device and then manipulated in any specific way must gain consent and may only use the data for an appropriate and identified purpose. Any additional processing is subject to strict control and possibly legal sanction.

CONCLUDING REMARKS AND RECOMMENDATIONS

[5] UK ICO, 'Personal information online code of practice' (July 2010) <http://www.ico.org.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/personal_information_online_cop.pdf> accessed 15th July 2013.
[6] DPD (n3) Article 6(1).
[7] By contrast to a *natural person*, a *legal person* is an entity that has acquired the status of a person through law. This is useful for example as it enables companies to enter into contracts. In a business to consumer contract, the business will be a legal person

Users	There is sufficient legislation in place to protect personal data. Data collection should be with consent, appropriate to the intended processing, and protected from inaccuracy as well as disclosure[8].
Providers	The Data Protection Framework would class application and service providers either as data controllers or data processors. Either way, they are expected to act according to their responsibilities in respect of personal data. Specifically, they should seek consent from the data subjects, as well as retain and processed data only within the limits of the original purpose.
FI Service and Application Developers	1) Be clear on your rôle as data processor or controller; and make users aware 2) Ensure explicit consent is requested for any data processing undertaken 3) Treat any consent forms in much the same way as a provider might manage SLAs: there should be dynamic compliance checking to the terms agreed to under data subject consent as an integral part of the execution of the application or service you run.

The reality: *should we be nervous?*

Issue	Within an FI ecosystem of competing requirements and drivers, the question is whether the legislation is adequate and whether indeed the authorities are willing to prosecute transgression. What is more, is it enough to expect users to agree to long-winded and complicated terms and policies thereby relieving developers and providers of the burden of appropriate data handling?

The most recent case for Google sees end users under siege again: *those who correspond with Gmail users* should "have no expectation of privacy"[9]. The end user licence agreement (EULA) presented to Gmail account holders bounds their own rights to privacy in what they send out. Invoking the "third party doctrine" for non-Gmail users, though, really should be challenged[10], not least on the basis that Google's intrusions are so extensive that implied consent is probably not enough.

However, there is a precedent for challenging such terms in the courts, even for Gmail users themselves (see below and the probation imposed on *Twitter*). Notwithstanding specific differences across different jurisdictions, there is generally sufficient protection for individuals. Beyond the legal background, though, most providers will request some kind of consent via a licence. This presupposes that users look at the terms and conditions and the privacy policies of the particular service they wish to use. For common utilities[11], their terms and policies may be summarised as follows:

[8] Although legislation does exist to protect user data, it is far from unusual for providers to fall foul of such legislation.
[9] http://techland.time.com/2013/08/14/google-says-gmail-users-have-no-legitimate-expectation-of-privacy/
[10] http://www.theverge.com/2013/8/14/4621474/yes-gmail-users-have-an-expectation-of-privacy
[11] We looked at *Amazon Web Services, Doodle, Dropbox, Evernote, Google, GoToMeeting, Podio, Prezi, Scribd, SkyDrive, Skype, SlideShare*.

(a) None of them claimed ownership over content submitted to their services, but all required some form of licence to be granted.

(b) None of the tools or utilities impose terms which are particularly different to the others; however *Google* is arguably granted the most wide reaching rights through requiring rights to be granted to allow not only for operation as do other providers, but also for promotion, improvement and development.

(c) Some tools and utilities do not explicitly state which rights are granted and under what type of licence; most notably *Dropbox* and *Evernote*. This is possibly due to them wanting their terms to be seen as "user friendly" and absent of legalese[12].

(d) Generally speaking data deletion is up to the users: users should manually delete all data from the service/platform before cancelling. Users must also be aware of what data they share since this may still exist after cancellation of accounts.

(e) Whichever utility or tool is used, users are often required to check terms and policy changes regularly, since some providers will not alert them to any such changes.

Apart from (b), these terms in connection with the consent requested, it would seem reasonable to expect that data are therefore protected. This is not always the case, though. A 2012 study[13] revealed some disconcerting cases of a number of different service and application providers:

UNAUTHORISED DATA COLLECTION

Apple It was found that *iPhones* had secret files tracking location without the owner's permission or indeed knowledge[14]. They denied any wrong-doing, insisting that this was for legitimate purposes to help identify subsets of locations within larger databases to benefit the individual user. They promised to encrypt the data and reduce time spent on the device in the future.

Carrier IQ[15] In 2011, *Carrier IQ* was found collecting data related to usage and so forth without subscribers' knowledge nor their ability to opt out.

Google During the collection of *Street View* images, the mobile camera cars "inadvertently" picked up data from unsecured wireless networks for some four years. *Google* claimed that this was an isolated occurrence involving a single engineer acting with company authorisation. This was contested and the Federal Communications Commission fined them some $25,000 and complained that *Google* obstructed investigations.

Intel In 1999, *Intel* had to disable a feature on the Pentium III chip after public outcry over what was suspected to be a "super cookie" that could effectively track the user's surfing activities indefinitely.

Path The photo sharing app, *Path*[16], was found to be uploading address books

[12] This arguably leaves some ambiguity that may be detrimental to the user.
[13] http://news.cnet.com/2300-1023_3-10012162.html
[14] news.cnet.com/"http://news.cnet.com/8301-13579_3-20055885-37.html
[15] http://www.carrieriq.com/

from subscribers' devices without permission.

UNAUTHORISED SHARING

AOL 1998: a customer service agent released personal information about a subscriber to the Navy about his sexual orientation which led to his discharge from the forces.
2006: the company published the search history of more than 650,000 users. Even though "anonymised", the specific search history of individuals could still be tracked.

facebook In 2011, the Federal Trade Commission (FTC) claimed that *facebook* were not even complying with its own rules on data sharing and access, promising to take appropriate steps in future to make things more transparent for subscribers[17]

Twitter Also in 2011, *Twitter* were put on probation for 20 years by the FTC and forbidden for:

> "misleading consumers about the extent to which it protects the security, privacy, and confidentiality of nonpublic consumer information, including the measures it takes to prevent unauthorized access to nonpublic information and honor the privacy choices made by consumers." (*op cit*)

UNEXPECTED VULNERABILITY

Microsoft In 1999, *Microsoft* had to black out *Hotmail* for some 12 hours after discovering that subscriber accounts could be accessed by anyone with a web browser.

Sony Protection software (a "rootkit") to avoid illegal copying installed itself when a CD was played, which rendered the machine vulnerable to malware. (2005)

UNEXPECTED SHARING

Yahoo! *Yahoo!* co-operated with the Chinese authority and released IDs of political dissidents who were subsequently imprisoned.

All of these cases, with the possible exception of *Yahoo!*, should be covered by the appropriate legislation: the data subjects should know what data are being collected, how they are being used, and who has access to them. Clearly, the terms and policies of providers cannot necessarily be taken as the final arbiter in such cases: both *Twitter* and *facebook* have been prosecuted for misleading terms or even failing to comply with their own terms.

CONCLUDING REMARKS AND RECOMMENDATIONS	
Users	Users need to review and maintain currency with the terms and conditions and privacy policies of the providers whose services they depend on.

[16] http://www.path.com
[17] http://ftc.gov/opa/2011/11/privacysettlement.shtm

Providers	Despite their terms and conditions, providers have been successfully prosecuted for a number of different failings, including unauthorised collection, use and disclosure of personal data, even deliberately over-complicating the terms of service they impose.
FI Service and Application Developers	1) Services should use specific terms and conditions appropriate to their purpose; terms should be simple to understand; users should be alerted to changes *and what that means to them* 2) Services should not involve hidden data collection 3) Services should not disclose or share data without explicit consent.

User perceptions: *trust*

Issue	For the FI to deliver on the promise of economic growth and expansion, as well as online activity to become the norm and available to all, there is a significant and fundamental question: will the end-users, the consumers of the services driving the *Digital Agenda's* virtuous cycle, be prepared to embrace the new technologies and engage whole-heartedly in that context? For the FI to succeed, users must trust that they will be dealt with appropriately.

So, there is legal protection under the Data Protection Directive and clear legal precedent that data misuse, as well as obviously misleading terms, will be challenged. That being said, the question is whether or not individuals do actually feel they can trust the services they depend on. The 2012 results[18] for the annual survey of some ten thousand adults into how they rate organisations in terms of their "[commitment] to protecting the privacy of their personal information" reveals some interesting trends. In summary, and ignoring what organisations were mentioned and indicators of rankings between them, the following results may be cited:

Attitudes to privacy	• 78% of respondents continue to see the protection of their personal information as instrumental in building and maintaining trust[19].
	• Important measures related to trust:
	  ○ 73% : security over personal information
	  ○ 59% : no data sharing without consent
	  ○ 59% : the ability to be forgotten [*sic*]
	  ○ 55% : the right to revoke consent.
	• 49% reported receiving one or more data breach notifications within the previous two years, of whom 70% said it reduced their level of trust.

Most trusted sectors	From 25 categories, *healthcare, consumer products* and *banking* are the most trusted in terms of privacy; *Internet* and *social media*, *charities* and *toys* are the

[18] http://www.ponemon.org/local/upload/file/2012%20MTC%20Report%20FINAL.pdf

[19] Nevertheless, 63% admit sharing sensitive personal information with organisations they didn't know or trust, of whom 60% justified it on the basis of convenience (i.e. making a purchase).

	least trusted.
Effects of technology	• 59% believe privacy rights undermined by disruptive technologies (social media, smart mobile devices, geo-tracking tools) • 55% claim that privacy has been diminished because of government intrusions
User control	• 35% believe they maintain control of their own information. This figure has been going down for some seven years. • 61% (the highest) believe *identity* is the main concern related to privacy, while • 56% cite an increase in government surveillance.
View of policies	32% do not rely on policies when making trust judgements, of whom 60% claim the policies are too long or contain too much legalese.

Irrespective of individual conclusions, perhaps the most interesting results in the present context suggest that social media and the Internet in particular lack trust. This clearly does not stop subscribers using them[20], even though long-term engagement via SNS can be significantly less satisfying than direct social interaction[21]; nor institutions and agencies forcing contact online[22]. Looking further, though, it is interesting to note:

1) Users are concerned about a *loss of control*, and yet the DPD and the explicit requirement for consent associated with specific data handling, should provide such control; and
2) Users bemoan *government intrusion*[23]. However, this is treated as an exception in the DPD[24],[25].

Dutton and his colleagues have noted that Internet users with increasing experience and familiarity develop appropriate trust levels online, and decide for themselves what they can and should not do: that is, irrespective of the regulatory framework, experienced users will make their own decisions

[20] The current top three include *facebook* (1,000M subscribers), *Twitter* (500M) and *Google+* (500M). See http://news.cnet.com/8301-1023_3-57525797-93/facebook-hits-1-billion-active-user-milestone/; and also http://news.discovery.com/tech/apps/top-ten-social-networking-sites.htm on site popularity, and http://social-networking-websites-review.toptenreviews.com/ on site rankings, including interestingly "security".

[21] http://uk.news.yahoo.com/facebook-social-network-linked-unhappiness-215939039.html#m6L7M0L

[22] Such as the US immigration authority (http://travel.state.gov/visa/forms/forms_1342.html) ; the Digital Agenda for Europe is also seeking to encourage online participation (http://ec.europa.eu/digital-agenda/en/scoreboard, and especially Pillars IV, VI and VII).

[23] See recent discussion about NSA (eg http://www.theguardian.com/world/2013/jul/25/justice-department-case-nsa-collection) and the case of Prism (http://www.bbc.co.uk/news/technology-23051248)

[24] Article 13, for instance, provides for member states to circumvent requirements on data privacy on the basis of *national security* i.a.

[25] The disclosure of *Yahoo!* cited earlier was generally frowned upon in the US courts, yet highlights some level of ambiguity in government position on "surveillance" and overriding basic privacy rights. This was also seen in government attitudes to the Arab Spring *versus* the London Riots motivated by police shooting of Mark Duggan.

and judgements[26]. It is nonetheless true that users are, and should be, nervous about information being collected about them, especially now that the various technologies and services they use readily allow data to be collected and analysed[27]. The SESERV project (http://www.seserv.org) concluded with a specific recommendation on this point: service and application providers as well as government and other agencies should not "let the ease of collecting user data be done to such an extent as to let the user feel under surveillance or threat (or more simply put off)"[28]. The problem is exacerbated though by the fact that the data which are collected are not necessarily just personal information covered by the DPD: they may also include online activity (searches and so forth) which can easily be cross-correlated with the personal data available through SNS[29].

Finally, in this section, consider the provision of the DPD on retaining data. Article 6, section (iv) talks about data being kept "up to date" and goes on to stipulate that inaccurate or incomplete data should be erased or rectified (see above). From a user perspective, it would be tempting to assume that this actually means that as the data subjects they have the right to remove those data when they no longer want them[30]. However, as an Austrian student discovered, this is not necessarily the case, with *facebook* retaining all his personal data even those he thought he had deleted. This led to a request under the Freedom of Information act which revealed the extent of the problem[31].

Users do remain uncomfortable about privacy and the protection of their personal data, therefore. In particular, the capabilities of disruptive technologies (smart phones, SNS, etc.) as well as government surveillance rate high on the list of concerns. Despite legislation which outlines appropriate use and attempts to limit processing, as well as clear indications of a will to prosecute where necessary, users still complain of a loss of control and concerns over intrusion.

CONCLUDING REMARKS AND RECOMMENDATIONS	
Users	Users need to monitor their own use of sites and services if they feel under threat. Provision is made for protection of their private information, but they should be vigilant themselves.
Providers	Care needs to be taken to avoid a perception of snooping. Increased transparency about data handling would help build and maintain trust among users. Government agencies should be particularly mindful of user concerns.
FI Service and Application Developers	1) Make it easy for users to view their personal data, modify them and/or remove them; 2) Make sure that any such activity is applied across all data you hold; 3) Allow users to view before and after "states" (i.e. to alleviate fears of surveillance)

[26] Dutton, W.H. and Shepherd, A. (2003) "Trust in the Internet: The Social Dynamics of Experience Technology", The Oxford Internet Institute, available from: http://www.oii.ox.ac.uk/resources/publications/RR3.pdf

[27] See the summary results from the Ponemon survey above.

[28] Recommendation 9 http://www.scribd.com/doc/105908010/D3-1-2-v2-pdf

[29] Krishnamurthy, B. and C.Wills (2010) On the Leakage of Personally Identifiable Information Via Online Social Networks. ACM SIGCOMM Computer Communication Review, Volume 40, Number 1, pp. 112 – 117.

[30] See also the "ability to be forgotten" in the Ponemon trust survey.

[31] http://europe-v-facebook.org/EN/en.html

User confidence: *the public domain*

Issue	One of the major new technologies to emerge late in the 90s but increasingly in the first decade of this millennium is social media. End users are now able to interact with others – family, friends, associates, people with similar experiences and interests – via the Internet. Lives are increasingly lived out in part at least through social networks with *tweets* for example keeping anyone and everyone abreast of what is happening to, for and with anyone else. Constant sharing of personal experience in this way begs the question of whether such information and data should not be treated as public property the processing of which does not require regulation. When is my personal data no longer my property?

The SESERV recommendation referred to previously continues with:

> "[…] citizens have a responsibility to careful about what information they share online, and regulators need to educate the public about and enforce data protection laws."[32]

putting a logically founded onus on users to take on their fair share. If there are terms and policies, however long-winded and linguistically unfamiliar, they should be read; if users have concerns, they should raise them, and ask for disclosure of what is personal to them[33]. There is, however, another angle to this: in disclosing information about myself in an SNS, do I have a right to privacy or is that information now public?

The concept of the "public domain" has been called a "multifaceted and multidimensional concept with no definite definition",[34] indeed its remit changes depending on its subject. The concept itself may refer to:

1) public domain in relation to data and information and how this interacts with data protection; and

2) the broad area of intellectual property, which itself has different conceptions of public domain depending on the type of IPR.

In the present context, we will consider (1) above only since we are more concerned with personal data and its disclosure than issues around the protection of creativity associated with content generated by users.

European Union Level

At the EU level, the data protection framework provided for in the DPD does not distinguish between personal data that is in the public domain until that data falls under Article 8 as "special categories of data".

[32] *Loc cit*
[33] This may not always be straight-forward: going through all terms and policies online may be too time-consuming, or when wishing to use a service for the first time motivation may be low.
[34] Tshimanga Kongolo, 'Intellectual property and misappropriation of the public domain' (2011) 33(12) EIPR 780

Article 8(1) establishes that no processing of personal data which reveals "racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership" shall occur and nor shall "processing of data concerning health or sex life". This statement is then subject to certain exclusions listed in Article 8(2) as

> "(a) the data subject has given his explicit consent to the processing of those data, except where the laws of the Member State provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject's giving his consent; or
>
> (b) processing is necessary for the purposes of carrying out the obligations and specific rights of the controller in the field of employment law in so far as it is authorized by national law providing for adequate safeguards; or
>
> (c) processing is necessary to protect the vital interests of the data subject or of another person where the data subject is physically or legally incapable of giving his consent; or
>
> (d) processing is carried out in the course of its legitimate activities with appropriate guarantees by a foundation, association or any other non-profit-seeking body with a political, philosophical, religious or trade-union aim and on condition that the processing relates solely to the members of the body or to persons who have regular contact with it in connection with its purposes and that the data are not disclosed to a third party without the consent of the data subjects; or
>
> (e) the processing relates to data which are *manifestly made public by the data subject* or is necessary for the establishment, exercise or defence of legal claims." [our italics]

This is the only part of the Directive which distinguishes between data not in the public domain and data which is. There is no explicit EU law on what "manifestly made public by the data subject" means, but it is generally understood as requiring "a deliberate act by the data subject, disclosing the data to the public". For example, therefore, video surveillance would not be considered a conscious action to disclose information, but an interview to media, or publication on a public internet page would make data public.[35]

Following the absence of discussion regarding public domain throughout the rest of the Directive, it is clear that where personal data exists in the public domain and is subsequently processed, ***this processing is still subject to the Directive's principles and requirements*** [our emphasis].

Indeed, if one looks to the *UK ICO*, they noted in their online code of practice[36] that:

- "People may post their personal details in such a way that they become publicly visible – for example through a social networking or recruitment site. *Wherever the personal data originates, you still have an overarching duty to handle it fairly and to comply with the rules of data protection…*
- If you collect information from the internet and use it in a way that's unfair or breaches the other data protection principles, you could still be subject to enforcement action under the DPA even though the information was obtained from a publicly available source.

[35] Kotschy in Alfred Büllesbach et. al., *Concise European IT Law,* (2010, 2nd Edn, Kluwer Law International) 62
[36] UK ICO, 'Personal information online code of practice' (July 2010) <http://www.ico.org.uk/~/media/documents/library/Data_Protection/Detailed_specialist_guides/personal_information_online_cop.pdf> accessed 15th July 2013.

- It is good practice to only use publicly available information in a way that is unlikely to cause embarrassment, distress or anxiety to the individual concerned. You should only use their information in a way they are likely to expect and to be comfortable with. *If in doubt about this, and you are unable to ask permission*, you should not collect their information in the first place." [our italics]

Thus, if personal data is being processed that was collected from the public domain it must adhere to the data protection principles.[37] That is the data must be:

> "(a) processed fairly and lawfully;
>
> (b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes…
>
> (c) adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed;
>
> (d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified;
>
> (e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed…"

In essence, therefore, the final point in the UK ICO guidance is most important of the three: that the use must be "unlikely to cause embarrassment, distress or anxiety to the individual concerned" and a use of "their information in a way they are likely to expect and to be comfortable with". However, that said, all should be taken into account when considering the processing of publicly available data. This guidance implies an element of consent is granted through a data subject making that data publicly available, making the processing "lawful", and the controller must then ensure that the use does not cause embarrassment, distress or anxiety which meets the "fairness" requirement.

Individual jurisdictions vary in their interpretation of what constitutes "public domain", if at all. However, from a UK perspective, case law has shown that the determination of "public domain" falls to the question of whether the particular information was "'realistically' accessible to members of the public or only 'in theory'".[38] Further, information would not be in the public domain if it required an unrealistic "specialised knowledge and persistence" to find it. So the average member of public must be able to find the information fairly easily.[39]

The legislation in this area is not as clear cut. Nevertheless, and in summary, individuals who take steps to publish information in publically accessible areas which requires no significant effort to view, then they may be assumed to have released it into the public domain. That being said, there is still an obligation on others to process such personal information with care, specifically to avoid embarrassment and not in a way that would be unexpected by the individual concerned.

Anecdotal cases of embarrassment caused by making public responses best kept within the close circle of family and friends appear with continued regularity, including:

[37] DPD Article 6(1).
[38] Mosley v News Group Newspapers Ltd [2008] EWHC 687 (QB)
[39] Attorney General v Greater Manchester Newspapers [2001] EWHC QB 451

- Political gesturing[40]
- Social insensitivity[41]
- The long-term effects of past indiscretion[42]
- Unexpected consequences
    - of heroic intervention[43], or
    - public insensitivity[44].

But in all such cases, it could be argued that those posting such content should have had an expectation that they were making it public: these were "deliberate act[s] by the data subject, disclosing the data to the public".

There are positives and negatives here. Richard Branson, for instance, the importance of going public in this way:

> "Embracing social media isn't just a bit of fun, it is a vital way to communicate, keep your ear to the ground and improve your business"[45]

though this does have a knock on effect. Take employment, for instance:

- prospective candidates are encouraged to avoid specific types of activity (inappropriate comments or photos, dishonesty, and so forth)[46]
- employers *do* use sites to screen candidates[47]
- *facebook* is particularly important, it appears[48]

and so on. Indeed the non-use of social media may be taken as suspicious[49] . So online activity is now part of individual life and will presumably continue to be so[50]. This does not, however, mean that all personal details posted on such sights can be used arbitrarily, even though they may be regarded as in the public domain. Any derivative processing may only be done within the constraints of the DPD, and should certainly not cause embarrassment or go beyond what might be expected by the data subject themselves.

CONCLUDING REMARKS AND RECOMMENDATIONS	
Users	There is no reason not to use social media to share content, and indeed it may be a necessary or expected part of everyday life. Notwithstanding any specific settings or other facilities offered by the site in question, that content becomes public domain. Users must take responsibility for what they share in this way; authorities (and providers) should really help users

[40] http://www.dailymail.co.uk/news/article-2082527/Diane-Abbott-Twitter-race-row-MP-faces-calls-resign-racist-tweet.html
[41] http://www.bbc.co.uk/news/uk-england-essex-23164829
[42] http://uk.news.yahoo.com/teen-crime-commissioner-offensive-tweet-row-081528626.html#Tzh3SvT
[43] http://news.sky.com/story/1064049/shark-wrestler-grandad-disgusted-by-sacking
[44] http://www.bbc.co.uk/news/uk-england-coventry-warwickshire-11068063
[45] http://www.linkedin.com/today/post/article/20121019130632-204068115-why-aren-t-more-business-leaders-online
[46] http://blog.reppler.com/2012/06/
[47] http://blog.reppler.com/2012/07/
[48] http://blog.reppler.com/2012/03/13/can-your-facebook-profile-predict-job-performance/
[49] http://blog.reppler.com/2012/08/
[50] See also: http://www.seserv.org/Studying-the-Future-Internet/doesyourbosstweet

Providers	understand the implications of posting content.
	Users sharing content and personal information, for instance via a SNS, should still be respected under the DPD. Their content and personal information should *not* be shared, disclosed, or otherwise processed if it might cause embarrassment or alternatively in a way that the data subject could not reasonably have expected.
FI Service and Application Developers	1) Alert users to the scope of sharing when they post content and personal information (i.e. extend the *Are you sure?* type warnings to include *You are now releasing this information to these people*). 2) Restrict access to personal data: a. 3rd parties can only *view* but not *copy* b. APIs should disallow extraction without alert to data subjects c. In-house analytics should not be used[51].

User profiles and data mining: *derivative works*

Issue	With the technologies of the FI making ever more data available and end users increasingly willing or required to share personal data online, the next major challenge for the FI is the regulation of how such data is processed and interrogated to reveal even more about the lives and preferences of those online than they themselves might have wanted to share. What will the status be if we let the power of data analytics loose on data otherwise shared for different purposes?

So far in this section, we have considered issues about privacy around personal data, user perceptions of how much protection they are given even when they choose themselves to make the data public. Finally, we should consider the additional and perhaps more recent issue of how those personal data are manipulated to generate a derivative set of "meta data". As highlighted previously, it is not difficult to match online activity with personal details[29] and thereby potentially gain access to much more information than the data subject might have intended or expected. Irrespective of individual consent, there are two particular problems here:

1) In consenting to the limited use of personal data in different contexts, albeit in some cases embedded with the terms and policies of individual platforms, the user may not have intended the data to appear or be used together;
2) And more worryingly, data mining may be used to infer things which are not in fact true, or certainly would not have been released to anyone by the individual data subject.

In the first case, the data subject retains some level of control: they are still able to give or refuse consent for certain types of information or personal data. Even so, there is little protect against the

[51] Further processing in this way may well go beyond reasonable and expected additional processing, and as such may well go against the original terms of the consent provided. More important and potentially disturbing, though, is that further analyses may reveal hidden or unknown characteristics of users which they would certainly never have agreed to release.

possibility of jigsaw attacks[52] where a so-called motivated intruder might use legitimately available data sources and match one against the other. There is nothing that can be done against this, unless previously collected personal data have been successfully anonymised prior to release and in order to reduce the possibility of other datasets being linked to it.

The second case is perhaps more disturbing. Take the case of a retailer who simply through the analysis of sales established who may or may not be pregnant and targeted marketing material at those customers. When one of them turned out to be a teenage girl who had not yet decided to tell her family of her pregnancy, the whole family suffered[53]. There is nothing really in the legislation to protect against this, assuming she had released her name and address during her purchasing and had agreed to the information being used for marketing purposes, a fairly common request especially for online services and retailers.

The basic business model of SNS relies on precisely this kind of analysis: using the personal data that subscribers provide, along with monitoring their activities – their *Likes* and *Dislikes*, who they interact with, and the topics that motivate their participation – the site can provide targeted marketing and charge premium rates to do so. It is one thing to be irritated by such marketing, but quite another to discover prospective employers or even government agencies, such as the police and intelligence services, can do the same sort of analysis if they so desire. Innocent information from photos, for instance, including not only the subscriber or other members of the social networking site but also other people, could be used to identify the movements of individuals who had never provider consent in the sites end-user licence agreement. As such, data mining in this way could have a very serious and detrimental effect. On the one hand, it could reveal information that the individual would never have disclosed of their own volition. More disturbingly, if the analysis is incorrect, it will reveal information about an individual which isn't even correct[54],[55].

In some sense, the creation of additional metadata about an individual is legitimate, if ethically questionable: analysis could be covered under a usage agreement. In addition, the personal data used could be said to be in the public domain as previously discussed. However, and as outlined in the DPD, due care has to be exercised where embarrassment might result, and where the user would not have expected such additional analyses to take place. Significantly, though, perhaps it could be argued that information about pregnancy, sexual orientation or health which have clearly emerged and been disclosed in this way come under the category *sensitive data* and therefore may not be released without the explicit consent associated with them from the data subject.

CONCLUDING REMARKS AND RECOMMENDATIONS	
Users	The encouragement to be careful about what individuals reveal about

[52] See http://www.ico.org.uk/~/media/documents/library/Corporate/Research_and_reports/anonymisation_cop_draft_consultation.ashx and https://www.ico.org.uk/~/media/documents/library/Data_Protection/Practical_application/anonymisation_code.ashx

[53] Kashmir Hill, 2012. "How Target figured out a teen girl was pregnant before her father did," *Forbes* (16 February), at http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/, cited in Obler, Welsh and Cruz "The danger of big data: Social media as computational social science" http://firstmonday.org/ojs/index.php/fm/article/view/3993/3269

[54] Current accuracy rates may be no better than 70% http://www.theguardian.com/news/datablog/2013/jun/10/social-media-analytics-sentiment-analysis

[55] http://cacm.acm.org/magazines/2013/5/163753-discrimination-in-online-ad-delivery/fulltext provides a disturbing example of the use of derivative analyses.

	themselves is more acute when we consider: i. Data from different sources may be able to be cross-matched and reveal more about the users than they originally and separately intended; ii. Metadata can be derived from personal data along with online activity that may reveal information about the user that may cause embarrassment or worse, whether or not it is correct. There is also no clear legal protection against this.
Providers	Irrespective of consent and usage licences, providers have an ethical obligation to be sensitive about the data they derive from an original source. Care should be taken to protect the individual even in the face of market pressures.
FI Service and Application Developers	1) Despite the lack of clarity on the legal status of derived metadata, care should be taken if those data could be construed as *sensitive data*. 2) Disclosure of personal data even with consent should be done on a case-by-case basis and in consideration of the possibility of jigsaw attacks. 3) Users should be allowed to review and modify any derived metadata as part of the administrative components of any ongoing service or application.