Position statement

Yi Yin
Faculty of Technology Policy and Management
Delft University of Technology
Jaffalaan 5, 2628BX, Delft, The Netherlands
y.yin@tudelft.nl

My background in Privacy

I'm a PhD candidate from the ICT research group of the Faculty of Technology, Policy and Management at Delft University of Technology. My research focuses on data-driven privacy and trust enhancement mechanisms, specifically designing privacy and trust enhancement mechanisms for research data sharing. I'm also involved in the EU H2020 research project **VRE4EIC (www.vre4eic.eu)** which aims to build a multi-disciplinary virtual research environment to integrate various research datasets and orchestrate various data related web services built on top of EU research infrastructures.

When we integrate massive amounts of research data, we face many privacy challenges while we need fast data governance and approval of data sharing in the increasingly collaborative research activities. Traditional privacy enhancement mechanisms derived from information security practice are not suitable anymore for the semantic web. Within VRE4EIC, I'm involved in the work group working on handling security, privacy and trust issues related to research data sharing. We have investigated and published many solutions in these aspects.

Potential for improvement of interoperability

Researchers are concerned with potential privacy issues when sharing their research data with researchers from other organizations and other countries. Research data often require removing some privacy sensitive attributes before publication or sharing with other parties, since such attributes may lead to the identification of individual persons in these datasets . However, at the same time it is often not clear which attributes need to be removed exactly or how these attributes should be removed. There is no complete list of attributes that are assessed as privacy sensitive, since the privacy sensitivity of these attributes also depends on their combinations and the context in which they are used. Moreover, the combination of several datasets from different sources make it possible to re-identify an individual person or a small group of persons, especially when open linked data are combined with social media data. When several datasets are combined, it is not clear whether the attributes should be horizontally or vertically removed. Some datasets cannot be published in an open environment but require a more secure space for usage, or different levels of openness. Privacy and data protection legislation prescribe how one should deal with privacy information from abstract level, while many researchers are not even aware of the new GDPR. There is a large group of researchers who might unconsciously disclose research data which contain sensitive privacy information. Researchers may not always know under which conditions they are (legally) allowed to share their data (with different levels of sensitivity), with whom they can share their data, under which conditions they are allowed to reuse data from other researchers. Guidelines need to be given to the researchers to give sufficient space for the interpretation of privacy sensitivity. Because researchers often lack legal

expertise, it would make researchers' research life much simpler and easier if an automatic semantic privacy labelling and policy setting services on the basis of semantic information management model can help them define the *privacy sensitivity levels*, and *permissions*, *prohibitions* and *duties* related to the usage of data asset before they want to share their research data assets.

Potential for improvement on policies based information governance

After analysing many taxonomies in the literature related to privacy enhancement research, I found that most PETs explore the technical solutions derived from classical information security approach, e.g. focusing on anonymized access or cryptographic algorithms. The concept of information privacy and information security create confusion, although they are quite related. Information privacy is often perceived as a part of information security. In most situation, the goals of information privacy and security are the same or orthogonal. Therefore, privacy needs to be preserved or enhanced with the support of information security mechanisms. However, they can be also in conflict. For example, a social network system is a place where privacy and security can conflict. Facebook, Twitter, and Instagram allow people to communicate with each other without letting others know your real identity while people want the security of really knowing with whom they are exchanging information. In this situation, most systems are designed with functionalities for users to self-determine whether their privacy information is available in the communications. Another example is that a security system can require users with authenticated credentials to access a network without restricting access to personal information. This system can have security but not privacy. Different context makes the privacy sensitivity rather complex. We need data governance policies enabled by semantic information management models to clearly define the access policies in different contexts and integrate them through privacy by design.