Data protection and database theory - Applying database design principles to personal data identification

Dr Boštjan Berčič
Institute for legal informatics, Ljubljana, Slovenia
Email: bostjan.bercic@ipri-zavod.si
Dr Carlisle George
School of Computing Science, Middlesex University
Email: c.george@mdx.ac.uk

Abstract

The European Union (EU) directive on personal data[1], and resulting data protection legislation of EU member states, require from data controllers, a notification of their activities to the appropriate supervisory authority[2]. Included in this notification is also a description of the data or categories of data which are processed[3]. Legislation in some EU member states (e.g. Slovenia) require that not only a description but also a concrete list of personal data attributes need to be included in this notification. In such cases, it is sometimes difficult to ascertain *in concreto* whether some collected attribute represents personal data (and should therefore be included in the list of attributes) or whether it is a non-personal attribute. Similarly, under the EU directive data, subjects have various rights, including the right to access their data, and data controllers are sometimes faced with the problem of determining whether various data items constitute personal data.

According to the EU directive, personal data is 'any information relating to an identified or identifiable natural person', whereas 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'. In short, any information about an individual and his relations with other individuals and things. In the UK, however, in the case of *Durant v Financial Services Authority (2003 EWCA Civ 1746)*, the Court of Appeal issued a landmark ruling narrowing the interpretation of what makes data 'personal' (within the meaning of personal data under the EU directive and UK Data Protection Act 1998). The Court ruled that personal data is information which: is biographical in a significant sense; has to have the individual as its focus; and has to affect an individual's privacy whether in his personal family life, business or professional activity. This ruling in effect, narrows the right of subject assess under the EU directive and UK Data Protection Act 1998. It is therefore currently being taken before the European Court of Human Rights as a breach of Article Eight of the European Convention of Human Rights[4]. The European Court's decision, however, will take several years, but in the meantime the uncertainty about 'personal data' remains.

In view of the legal uncertainty regarding what constitutes personal data, this paper examines whether database design principles can be applied to personal data identification. Using this novel approach, the paper explores various parallels between personal data identification and principles of database design. For example, in attempting to answer questions such as what constitutes a personal data identifier (which indicates the presence of personal data) and which of the data items related to a personal data identifier are also personal data, the paper uses principles of database design. It applies the theory of database keys[5] to distinguish data that is 'personal' from that which is 'non-personal' and applies principles of database normalization (such as removing functional dependencies) to distinguish personal data attributes from other attributes merely connected in some way to personal data. Using numerous examples, the paper discusses (and demonstrates) how knowledge of database design principles can greatly help to understand what is and what is not personal data. Consequently, data controllers can make use of this in their notification processes as required under EU data protection legislation. Data controllers can also use this approach to determine 'personal data' when complying with the right of subject access (subject to other provisions under the Directive such as the rights of third parties). The paper thus makes a novel contribution to the ongoing uncertainty in data protection law, by the development of a decision algorithm to determine the presence of personal data. The

paper also discusses the wider issue of applying computing/scientific principles to interpreting the law, and comments on the success of the approach taken.

---

[1] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data
[2] For example The UK Information Commissioner's Office
[3] Article 19 (1)(c) of Directive 95/46/EC
[4] Article Eight of the European Convention states that everyone has the right to respect to his private and family life, his home and his correspondence.
[5] A key is a minimal set of attributes whose values uniquely identify an entity in a set.