This is the pre-print version of a paper accepted in *Phil. Trans. R. Soc. A* 374: 20160112.

http://dx.doi.org/10.1098/rsta.2016.0360

What is Data Ethics?

Luciano Floridi[1,2] and Mariarosaria Taddeo[1,2]

[1]Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford, OX1 3JS, United Kingdom; [2]Alan Turing Institute, 96 Euston Road, London, NW1 2DB, United Kingdom.

Corresponding author: luciano.floridi@oii.ox.ac.uk

Abstract

This theme issue has the founding ambition of landscaping *Data Ethics* as a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including AI, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values). Data Ethics builds on the foundation provided by Computer and Information Ethics but, at the same time, it refines the approach endorsed so far in this research field, by shifting the Level of Abstraction of ethical enquiries, from being information-centric to being data-centric. This shift brings into focus the different moral dimensions of all kinds of data, even the data that never translate directly into information but can be used to support actions or generate behaviours, for example. It highlights the need for ethical analyses to concentrate on the content and nature of computational operations—the interactions among hardware, software, and data—rather than on the variety of digital technologies that enables them. And it emphasises the complexity of the ethical challenges posed by Data Science. Because of such complexity, Data Ethics should be developed from the start as a macroethics, that is, as an overall framework that avoids narrow, *ad hoc* approaches and addresses the ethical impact and implications of Data Science and its applications within a consistent, holistic, and inclusive framework. Only as a macroethics Data Ethics will provide the solutions that can maximise the value of Data Science for our societies, for all of us, and for our environments.

Keywords: Data Ethics, Data Science, Ethics of Data, Ethics of Algorithms, Ethics of Practises, Levels of Abstraction.

Data Science provides huge opportunities to improve private and public life, as well as our environments (consider the development of smart cities or the problems caused by carbon emission). Unfortunately, such opportunities are also coupled to significant ethical challenges. The extensive use of increasingly more data—often personal if not sensitive (Big Data)—the growing reliance on algorithms to analyse them in order to shape choices and to make decisions (including machine learning, AI, and robotics), as well as the gradual reduction of human involvement or even oversight over many automatic processes pose pressing issues of fairness, responsibility, and respect of human rights, among others.

These ethical challenges can be addressed successfully. Fostering the development and applications of Data Science while ensuring the respect of human rights and of the values shaping open, pluralistic, and tolerant information societies is a great opportunity of which we can and must take advantage. Striking such a robust balance will not be an easy or simple task. But the alternative, failing to advance both the ethics and the science of data, would have regrettable consequences. On the one hand, overlooking ethical issues may prompt negative impact and social rejection, as it was the case, for example, of the NHS care.data programme.[1] *Social acceptability* or, even better, *social preferability* must be the guiding principles for any Data Science project with even a remote impact on human life, to ensure that opportunities will not be missed. On the other hand, overemphasizing the protection of individual rights in the wrong contexts may lead to regulations that are too rigid, and this in turn can cripple the chances to harness the social value of Data Science. The LIBE amendments initially proposed to the European Data Protection Regulation offer a concrete example of the case in point.[2]

Navigating between the Scylla of social rejection and the Charybdis of legal prohibition in order to reach solutions that maximise the ethical value of Data Science to benefit our societies, all of us, and our environments is the demanding task of *Data Ethics*. In achieving this task, Data Ethics can build on the foundation provided by Computer and Information Ethics, which has focused for the past thirty years on the main challenges posed by digital technologies (Floridi 2013; Bynum 2015; Miller and Taddeo 2017). This rich legacy is most valuable. It also fruitfully grafts Data Ethics onto the great tradition of ethics more generally. At the same time, Data Ethics refines the approach endorsed so far in Computer and Information Ethics, as it changes the Levels

[1] See https://www.england.nhs.uk/ourwork/tsd/care-data/

[2] Amendments 27, 327, 328, and 334-3367 proposed in the Albrecht's Draft Report, http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/pr/922/922387/922387en.pdf.

of Abstraction (LoA) of ethical enquiries from an information-centric (LoA_I) to a data-centric one (LoA_D).[3]

Ethical analyses are developed at a variety of LoAs. The shift from LoA_I to LoA_D is the latest in a series of changes that has characterised the evolution of Computer and Information Ethics. Research in this field first endorsed a human-centric LoA (Parker 1968), which addressed the ethical problems posed by the dissemination of computers in terms of professional responsibilities of both their designers and users. The LoA then shifted to a computer-centric one (LoA_C) in the mid 1980s (Moor 1985), and it changed again at the beginning of the second millennium to LoA_I (Floridi 2006).

These changes responded to rapid, widespread, and profound technological transformations. And they had important conceptual implications. For example, LoA_C highlighted the nature of computers as universal and malleable tools. It made it easier to understand the impact that computers could have on shaping social dynamics as well as on the design of the environment surrounding us (Moor 1985). LoA_I then shifted the focus from the technological means to the content (information) that can be created, recorded, processed, and shared through such means. In doing so, LoA_I emphasised the different moral dimensions of information—i.e., information as the source, the result, or the target of moral actions—and led to the design of a macroethical approach able to address the whole cycle of information creation, sharing, storage, protection, usage, and possible destruction (Floridi 2006).

Data Science, as the latest phase of the information revolution, is now prompting a further change in the LoA at which our ethical analysis can be developed most fruitfully. In a few decades, we have come to understand that it is not a specific technology (computers, tablets, mobile phones, online platforms, cloud computing and so forth), but what any digital technology manipulates that represents the correct focus of our ethical strategies. The shift from *information* ethics to *data* ethics is probably more semantic than conceptual, but it does highlight the need to concentrate on what is being handled as the true invariant of our concerns. This is why labels such as "robo-ethics" or "machine ethics" miss the point, anachronistically stepping back to a time when

[3] The method of abstraction is a common methodology in Computer Science (Hoare 1972) and in Philosophy and Ethics of Information (Floridi 2008). It specifies the different LoAs at which a system can be analysed, by focusing on different aspects, called observables. The choice of the observables depends on the purpose of the analysis and determines the choice of LoA. Any given system can be analysed at different LoAs. For example, an engineer interested in maximising the aerodynamics of a car may focus upon the shape of its parts, their weight, and the materials. A customer interested in the aesthetics of the same car may focus on its colour and on the overall look and may disregard the shape, weights, and material of the car components.

"computer ethics" seemed to provide the right perspective. It is not the hardware that causes ethical problems, it is what the hardware does with the software and the data that represents the source of our new difficulties. LoA_D brings into focus the different moral dimensions of data. In doing so, it highlights the fact that, before concerning information, ethical problems such as privacy, anonymity, transparency, trust, and responsibility concern data collection, curation, analysis, and use, and hence they are better understood at that level.

In light of this change of LoA, Data Ethics can be defined as the branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing, and use), algorithms (including AI, artificial agents, machine learning, and robots), and corresponding practices (including responsible innovation, programming, hacking, and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values). This means that the ethical challenges posed by Data Science can be mapped within the conceptual space delineated by three axes of research: the ethics of data, the ethics of algorithms, and the ethics of practises.

The ethics of data focuses on ethical problems posed by the collection and analysis of large dataset and on issues ranging from the use of Big Data in biomedical research and social sciences (Mittelstadt and Floridi 2015), to profiling, advertising (Hildebrandt 2008), and data philanthropy (Kirkpatrick 2013; Taddeo forthcoming) as well as open data (Kitchin 2014). In this context, key issues concern possible re-identification of individuals through data-mining, -linking, -merging, and re-using of large datasets, as well as risks for so-called "group privacy", when the identification of types of individuals, independently of the de-identification of each of them, may lead to serious ethical problems, from group discrimination (e.g. ageism, ethnicism, sexism) to group-targeted forms of violence (Floridi 2014; Taylor, Floridi, and van der Sloot Forthcoming). Trust (Taddeo 2010; Taddeo and Floridi 2011) and transparency (Turilli and Floridi 2009) are also crucial topics in the ethics of data, in connection with an acknowledged lack of public awareness of the benefits, opportunities, risks, and challenges associated with Data Science (Drew forthcoming). For example, transparency is often advocated as one of the measures that may foster trust. However, it is unclear what information should be made transparent and to whom information should be disclosed.

The ethics of algorithms addresses issues posed by the increasing complexity and autonomy of algorithms broadly understood (e.g. including AI and artificial agents such as internet bots), especially in the case of machine learning applications. In this case, some crucial challenges include moral responsibility and accountability of both designers and data scientists with respect to unforeseen and undesired consequences as well as missed opportunities (Floridi 2012; Floridi forthcoming). Unsurprisingly, the ethical design and auditing (Goodman and Flaxman 2016) of algorithms' requirements and the assessment of potential, undesirable outcomes (e.g. discrimination or the promotion of anti-social content) is attracting increasing research.

Finally, the ethics of practices (including professional ethics and deontology) addresses the pressing questions concerning the responsibilities and liabilities of people and organisations in charge of data processes, strategies, and policies, including data scientists, with the goal to define an ethical framework to shape professional codes about responsible innovation, development, and usage, which may ensure ethical practises fostering both the progress of Data Science and the protection of the rights of individuals and groups (Leonelli forthcoming). Three issues are central in this line of analysis: consent, users privacy, and secondary use.

While they are distinct lines of research, the ethics of data, algorithms, and practices are obviously intertwined, and this is why it may be preferable to speak in terms of three axes defining a conceptual space within which ethical problems are like points identified by three values. Most of them do not lie on a single axis. For example, analyses focusing on data privacy will also address issues concerning consent and professional responsibilities. Likewise, ethical auditing of algorithms often implies analyses of the responsibilities of their designers, developers, users, and adopters. Data Ethics must address the whole conceptual space and hence all the three axes of research together, even if with different priorities and focus. And for this reason, Data Ethics needs to be developed from the start as a macroethics, that is, as an overall "geometry" of the ethical space that avoids narrow, *ad hoc* approaches but rather addresses the diverse set of ethical implications of Data Science within a consistent, holistic, and inclusive framework.

This theme issue represents a significant step in such a constructive direction. It collects fourteen contributions, each analysing a specific topic belonging to one of the three axes of research outlined above, while considering its implications for the other two. The articles included in this issue were initially presented at a workshop on "The Ethics of Data Science, The Landscape for the Alan Turing Institute" hosted at the

University of Oxford in December 2015. The issue shares with the workshop the founding ambition of landscaping Data Ethics as a new area of ethical enquiries and to identify the most pressing problems to solve and the most relevant lines of research to develop.

Before leaving the reader to the articles, we would like to express our gratitude to the authors and the reviewers for their contributions, as well as to the Alan Turing Institute for funding the landscaping workshop as part of its research strategy. The workshop and this special issue would not have been possible without the strong support and continuous encouragement of many colleagues, but in particular of Professor Helen Margetts, Director of the Oxford Internet Institute, and of Professor Andrew Blake, Director of the Alan Turing Institute. We are also grateful to Bailey Fallon, the journal's Commissioning Editor, and to the editorial office of *Philosophical Transaction A* for their great help during the process leading to the publication of this theme issue.

References

Bynum, Terrell. 2015. "Computer and Information Ethics." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2015. http://plato.stanford.edu/archives/win2015/entries/ethics-computer/.

Drew, Cat. Forthcoming. "Data Science Ethics in Governmen." *Philosophical Transactions A*.

Floridi, Luciano. Forthcoming. "Distributed Moral Responsibility in Networked Environments." *Philosophical Transactions A*.

———. 2006. "Information Ethics, Its Nature and Scope." *SIGCAS Comput. Soc.* 36 (3): 21–36. doi:10.1145/1195716.1195719.

———. 2008. "The Method of Levels of Abstraction." *Minds and Machines* 18 (3): 303–29. doi:10.1007/s11023-008-9113-7.

———. 2012. "Distributed Morality in an Information Society." *Science and Engineering Ethics* 19 (3): 727–43. doi:10.1007/s11948-012-9413-4.

———. 2013. *The Ethics of Information*. Oxford: Oxford University Press.

———. 2014. "Open Data, Data Protection, and Group Privacy." *Philosophy & Technology* 27 (1): 1–3. doi:10.1007/s13347-014-0157-8.

Goodman, Bryce, and Seth Flaxman. 2016. "European Union Regulations on Algorithmic Decision-Making and A 'right to Explanation.'" *arXiv:1606.08813 [Cs, Stat]*, June. http://arxiv.org/abs/1606.08813.

Hildebrandt, Mireille. 2008. "Profiling and the Identity of the European Citizen." In *Profiling the European Citizen*, edited by Mireille Hildebrandt and Serge Gutwirth, 303–43. Springer Netherlands. http://link.springer.com/chapter/10.1007/978-1-4020-6914-7_15.

Hoare, C. A. R. 1972. "Structured Programming." In , edited by O. J. Dahl, E. W. Dijkstra, and C. A. R. Hoare, 83–174. London, UK, UK: Academic Press Ltd. http://dl.acm.org/citation.cfm?id=1243380.1243382.

Kirkpatrick, Robert. 2013. "A New Type of Philanthropy: Donating Data." *Harvard Business Review*. March 21. https://hbr.org/2013/03/a-new-type-of-philanthropy-don.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. SAGE.

Leonelli, Sabina. forthcoming. "Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems." *Philosophical Transactions A*.

Miller, Keith, and Mariarosaria Taddeo, eds. 2017. *The Ethics of Information Technologies*. Library of Essays on the Ethics of Emerging Technologies. Abingdon, Oxon ; New York, NY: Routledge.

Mittelstadt, Brent Daniel, and Luciano Floridi. 2015. "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts." *Science and Engineering Ethics*, May. doi:10.1007/s11948-015-9652-2.

Moor, James H. 1985. "What Is Computer Ethics?*." *Metaphilosophy* 16 (4): 266–75. doi:10.1111/j.1467-9973.1985.tb00173.x.

Parker, Donn B. 1968. "Rules of Ethics in Information Processing." *Communications of the ACM* 11 (3): 198–201. doi:10.1145/362929.362987.

Taddeo, Mariarosaria. Forthcoming. "Data Philanthropy and The Design of The Infraethics for Information Societies." *Philosophical Transactions A*.

———. 2010. "Trust in Technology: A Distinctive and a Problematic Relation." *Knowledge, Technology & Policy* 23 (3–4): 283–86. doi:10.1007/s12130-010-9113-9.

Taddeo, Mariarosaria, and Luciano Floridi. 2011. "The Case for E-Trust." *Ethics and Information Technology* 13 (1): 1–3. doi:10.1007/s10676-010-9263-1.

Taylor, Linnet, Luciano Floridi, and Bart van der Sloot, eds. Forthcoming. *Group Privacy: New Challenges of Data Technologies*. Hildenberg: Philosophical Studies, Book Series, Springer.

Turilli, Matteo, and Luciano Floridi. 2009. "The Ethics of Information Transparency." *Ethics and Information Technology* 11 (2): 105–12. doi:10.1007/s10676-009-9187-9.

Vayena, Effy, and John Tasioulas. Forthcoming. "The Dynamics of Big Data and Human Rights: The Case of Scientific Research." *Philosophical Transactions A*.