



Report on ERC Exploratory Workshop “What Is Data-Intensive Science?” 17-19 December 2014, Exeter

26 January 2015, Niccolo Tempini & Sabina Leonelli

From Wednesday 17th to Friday 19th, December 2014, a group of philosophers and scientists convened at the Jury’s Inn hotel in Exeter to exchange ideas and perspectives, and assess the state of each other’s research interests and work, in regard to the epistemology of data-intensive science. The workshop was organized by Sabina Leonelli as a launch event for the ERC-funded project DATA_SCIENCE (www.datastudies.eu). It involved the research team Niccolo Tempini and Gregor Halfmann, the advisory board of the project and some additional collaborators with world-leading expertise on data-related research.

The aim of the workshop was to build on previously held discussions about the role of data and the forms of data practices in data-intensive sciences, so as to establish a research network around these issues. The workshop was inaugurated by a presentation by Sabina Leonelli on the current stage of development of her work on the philosophy of data, and on the ERC-funded project and its context. Among the overarching goals of the project, are investigating how data are handled across disciplines, and identifying and exploring cases of data re-use. The objectives of the workshop were to discuss relevant dimensions of data-intensive science, from empirically informed perspectives on specific target disciplines; to identify key research themes; to explore the potential for cross-disciplinary comparisons; and to discuss potential contributions by participants to an edited volume, provisionally titled “Comparative Perspectives on Data-Intensive Science”, to be assembled in 2017-18 as the output of these discussions.

During the first two days, the invited speakers shared their research interests on the topic and described the data projects in which they were involved. Talks came from a variety of disciplines (philosophy, history, social studies and natural sciences), and each talk focused on a specific target science. Disciplines covered included biology, astronomy, particle physics, meteorology, archaeology, zoology, biomedicine, history, economics, sociology, geography and science and technology studies. In the conclusive and third day two sessions were entirely dedicated to, first, recapitulating and commenting on the core themes that had emerged in the workshops and potential directions for comparison among disciplines and approaches, and second, strategising coordination towards the development of synergistic research in the years to come. Some initial plans for further collaboration and discussion, cutting across participants’ own disciplines of practice and of reference, started to be delineated.

In terms of topics, some lines of convergence came powerfully to the fore. The conversations started from ideas about salient **characteristics of data** (formats,

materiality, portability) and their complex relationships to **contexts**, which emerge in occasions of data generation, interpretation and inference but also management and placement. A starting point for the workshop was the idea of analyzing **data journeys**, including production, dissemination and re-use. This was favorably received, though it proved very hard to pinpoint similarities and differences among methods useful to 'follow the data' across different disciplines at this stage – and a crucial question raised by discussions concerned how 'stable' an object data can be, whether it has integrity or it is constantly mutable, and how does one follow data that keep shifting their format and significance across contexts. Much emphasis was placed on the ambiguous **status** of data as made and yet 'given', or 'raw' vs 'cooked'. Issues of **scale, extension and intensiveness** of data practices were discussed at length, e.g. the idea of 'big' data science, the different meanings of 'intensive' (use-intensive? Labour-intensive? Intensive relatively to other types of practices?) and the ways in which data acquire, shift and retain **value** (which strongly relate to various forms of ownership of data observed once data practices across fields are compared). Another important dimension of inquiry was found by all speakers to consist of the **relationship between data and other artifacts** for scientific research (e.g. specimens, equipment, infrastructures, metadata). Many speakers stressed the importance of various forms of regulation of **data access**, and the ethical issues involved. The question of how data are translated into useable, reliable, durable finds was raised several times, especially given existing conflicts and power struggles over what constitutes **secure data** and **valid interpretation**. Last but not least, and beyond the already innovative work concerning data **curation**, much attention was devoted to forms of **data labor** and to the differential involvement of social groups, identities and minorities in various forms of scientific work.

These discussions proved to us the fruitfulness of putting students of different scientific disciplines in dialogue with each other, and the importance of starting such conversations with a view to meeting again over the next few years and developing a common language and approach through which these themes can be framed and researched. We thus could not hope for a better launch event for DATA_SCIENCE.