Eight components for 'open social science' – An agenda for cultural change

The open science movement has been gathering force in STEM disciplines for many years, and some of its procedural elements have been adopted also by quantitative social scientists. However, little work has yet been done on exploring how more ambitious open science principles might be deployed across both the qualitative and quantitative social science disciplines. **Patrick Dunleavy** *sets out some initial ideas to foster a cultural shift towards open social science, explored in a current CIVICA project.*

The open science (OS) agenda has grown markedly in recent years, perhaps threatening to become 'all things to everybody' in the most inchoate accounts of what OS means. Yet so far, despite some optimistic recent findings, OS practices have made a relatively modest impression in many social science disciplines. Perhaps this reflects OS's overly 'physical' or STEM science roots. Some well-known open science websites seem to focus only on quantitative lab, experimental and computational work. Apparently excluded are even other STEM fields like ecology, where precise controls of background conditions for fieldwork are unfeasible. This pattern repeats in the social sciences, where some early handbooks and pioneer units in US universities promoting open social science ideas have focused solely on quantitative data analysis research, construed in narrow ways.

As part of a CIVICA research project funded by the EU's Horizon Europe programme, eight major European social science universities (LSE, Sciences Po, Hertie, Bocconi, EUI, CEU, Stockholm School of Economics, and NUPSPA in Romania) have come together to try to explore more broadly the authentic academic purposes that can be fostered by adopting open social science ways of working. We aim to go beyond the bureaucratic box-ticking character of some past, limited 'open science' approaches, and to make progress on a more ambitious academic agenda (even if some avenues may prove difficult). This piece covers initial thoughts on eight core principles, for which we are keen to hear comments and feedback.

Using open access publishing

Wherever feasible, research results and analyses should be published openly, making full text analyses and associated datasets maximally accessible to other researchers, including those in nearby disciplines. In addition, it is helpful if research results are also communicated to other potential audiences using modern, digital, and accessible social media. All modern social science work deals with 'human-dominated and human-influenced systems' that are fast-changing and complex. In this respect, the 'ordinary knowledge' of participants and practitioners, potentially engaged through open publication, has considerable value for contextualizing and improving scholarship (as Lindblom and Cohen famously argued). Effective dissemination also ensures the widest possible academic scrutiny of new work, improving fact checking, problem-solving, research accuracy and the emergence of diverse scholarly voices. The growth and improvement of 'citizen social science' also rely heavily on maximizing open access to scholarship. Finally, the experience of using open access publishing and impact-conscious dissemination can encourage researchers to delve further into other open science approaches.

## Making quantitative research datasets open and able to be reused

All the data supporting published quantitative research articles and books, plus details of their definitions, coding, and analyses, should be accessible – so that any of them can be easily checked if questions arise. In addition, researchers should take steps to encourage the wide re-use of their data by other analysts, conducting different kinds of studies. Most public and philanthropic grant-funders now require research data to be deposited in open archives (with well-explained documentation and data in accessible formats) once projects are completed. However, a great many corporate- or university-funded research datasets are not yet available. The growing importance of using proprietary social media data in research has further exacerbated this problem.

In other areas, the trend towards multi-national and multi-team research efforts has fostered the development of re-usable and comparable datasets covering multiple countries, or datasets that include pre-defined common questions that can be 'pooled' into more inclusive or authoritative resources. Achieving more improvements may entail new research taking better account of the best/strongest available research in a field (e.g., by using systematic reviews). Developing a culture of open data along these lines, would both promote and enable less silo-bound disciplinary approaches to defining questions and variables in quantitative social science.

Making quantitative research reproducible

The research results achieved by one team should be reproducible by other researchers using the same datasets and analysis methods (see fig.1). The endorsement of the well-known FAIR principles of data stewardship by many universities has extended this approach to cover smaller and self-funded studies. This is a wide imperative, one that goes far beyond the alleged 'reproduceability crisis' in some disciplines, such as psychology (often misleadingly termed a replication crisis). Many quantitative journals require a replication archive (RA) as a condition for publishing quantitative articles. Academic journals that have appointed specific Data Editors to check the quality, scope and useability of RAs have especially advanced a culture of openness.

Figure 1: The Turing Way diagram of reproducible, replicable, robust and generalizable research.
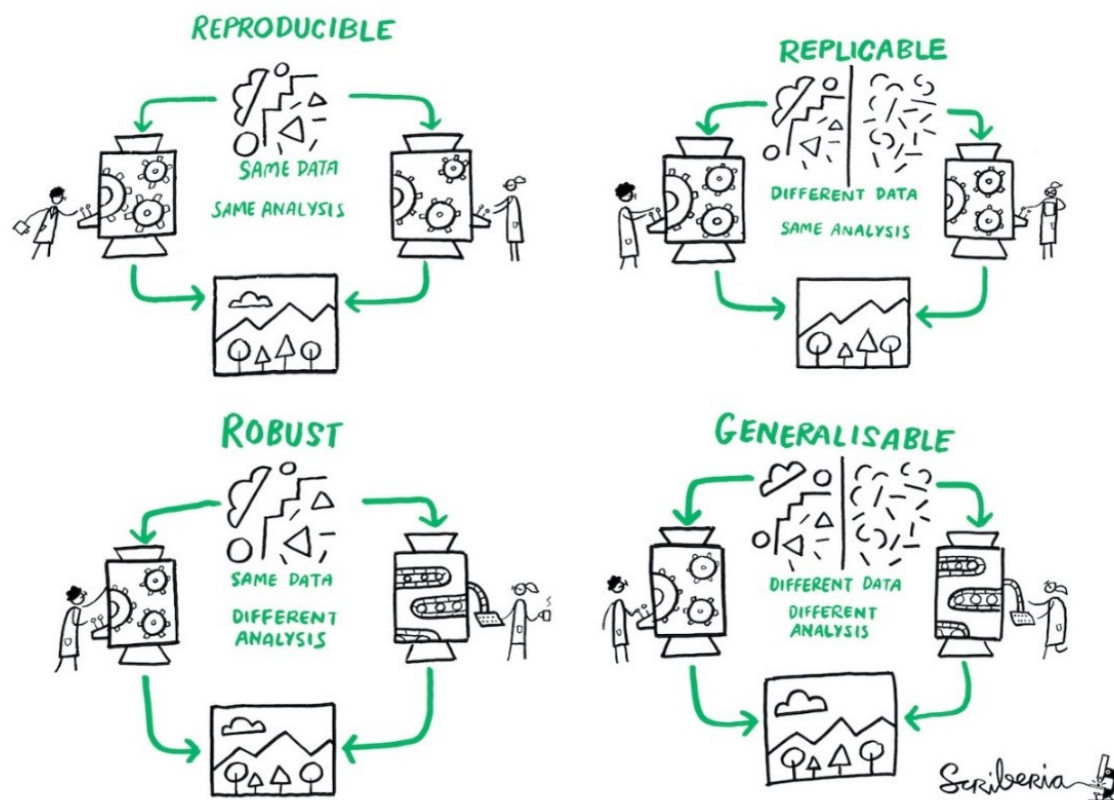
Source: The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 licence.
DOI: 10.5281/zenodo.3332807. See also here.

Yet improvements in reproducing results are hard to achieve, even in the experimental STEM sciences. Transferring methods between an original team and colleagues checking reproducibility entails overcoming tacit or implicit knowledge at many stages of a project. At a micro-level progress may rely on expanding the recording and documentation of information, data and analyses undertaken across multiple experiments, models, data-runs, fieldwork expeditions or archive visits. More systematic use of journaling, lab notebooks, and documentation of searches, coding analyses etc., are also potentially important here.

Improving the reproducibility and re-useability of qualitative evidence

Qualitative researchers in many social science disciplines have made far less progress in creatively adopting open science practices. There are undeniably [more logistical issues](#) in making knowledge, information and data drawn from interviews, archives, documentation, text sources, images and so on easily available for scrutiny or re-use, for diverse reasons – including information made available 'off the record' on 'non-attributably'; preserving the anonymity of research interviewees, patients, or those affected by problems being analysed; restrictions imposed by owner or museums on the reuse of images or scans of documents; or extensive documentary sources not being available in digital forms. We are still in the outermost foothills of exploring what open science may mean in qualitative research, and the scope of feasible changes is often contested. Yet the challenge is to find [new ways of undertaking research](#), and new steps that can mitigate the impact of current barriers.

Enhancing the replicability of research

As the Turing Way diagram above illustrates graphically, it is also vitally important for the integrity and prestige of all the social sciences that research analyses and results found in one dataset or body of evidence by one team can then be replicated by other researchers using the same treatment or methods on *different* data sources. Although no invariant 'natural world' lies behind the complexities in modern social sciences, similar situations and datasets should not produce radically divergent findings when analysed using the same methods – e.g., because the range of countries, dates, or data included changes somewhat. Improvements in replicability are especially difficult in field research disciplines like ecology and the non-experimental social sciences, where complex background conditions may be shaped by hundreds of causal influences, all operating at once and changing constantly.

Enhancing the robustness of research

When different teams of social scientists analyse the same datasets or bodies of evidence, using different analysis methods, they should none the less generate consistent results (Fig.1). Greater 'triangulation' of different research approaches focused on the same phenomena can generate stronger levels of assurance that results are not artefactual. In some areas of the social sciences the pooling and sifting of published evidence has been intensively studied (as with central bank economists seeking to incorporate the widest range of trends-data into modelling the whole economy), yet divergent overall judgements often still result. Improving the robustness of economic and social analyses is key if we are alter the widespread media assumptions that STEM scientists 'find' results, but social scientists can only 'argue' for their 'interpretation' of results.

Boosting the generalizability of research

We also ideally want studies of different sets or types of evidence gathered by different teams to none the less add up to a coherent, common picture of what is going on. Cumulating social science knowledge more could make it more applicable across a greater range of social settings and time periods, and allow research findings to be calibrated and adapted with some confidence to predict outcomes across diverse settings. Greater generalizability could also boost the confidence of both researchers and practitioners/citizens that academic research provides reliable bases for action, organizational decisions, or policy-making, and that some expert consensus exists.

Making more use of 'citizen social science'

Research results need to be communicated back more to the subjects, organizations and communities involved, and their reactions and understandings should be sought, assessed, and incorporated wherever feasible. The procedural steps in OSS outlined above create opportunities to keep participants informed or engaged throughout the process, rather than results only emerging months or years later with formal publication. This can be achieved via formal processes, where research informs participatory processes (as with deliberative democracy), where citizens and networks undertake some tasks themselves that feed into research, or where researchers and communities co-produce and co-steer outputs. However, taking citizen, user or practitioner feedback seriously can also be more generally practiced, in less formalized ways that none the less refine and strengthen researchers' understandings.

The sociologist Randall Collins famously argued that the STEM sciences succeeded so spectacularly in the last 150 years because they developed a 'rapid discovery /high consensus' approach to research. The greatest scientific prestige attaches to making new findings at the research frontier, while inside or behind the frontier 'normal science' levels of agreement are achieved on the prevalent scientific paradigm. By contrast, social science research has achieved far less standing because of deep theoretical/ideological dissensus on what counts as fundamentals, and a relatively high turnover of evanescent findings or 'fashions' of analysis that have lowered its prestige. Developing open social science is a movement that could re-focus social science towards achieving more *reliable* advances in knowledge at the research forefront, while also hopefully growing a somewhat greater (even if probably still moderate) level of social scientific consensus on the understanding of core societal processes.

_____

CIVICA Research brings together researchers from eight leading European universities in the social sciences to contribute knowledge and solutions to the world's most pressing challenges. The project aims to strengthen the research & innovation pillar of the European University alliance CIVICA. CIVICA Research is co-funded by the EU's Horizon 2020 research and innovation programme.

Note: This article was first published on the LSE Impact of Social Science blog.

About the Author

Patrick Dunleavy is Emeritus Professor of Political Science and Public Policy at the London School of Economics, and the Editor in Chief at LSE Press. With Helen Porter (LSE) and CIVICA colleagues, he leads a workgroup that is compiling an open science handbook for both the quantitative and qualitative social sciences. His most recent book is Maximizing the Impacts of Academic Research (Bloomsbury Press, 2020 (originally Palgrave), co-authored with Jane Tinkler.

Image credit: by Toni Reed on Unsplash.