

Panel 41. Sampling and the Making of Good Science: Examining Data Collection Practices and Their Implications

Convenors:

Victor Secco, Università Ca' Foscari Venezia

Valentina Marcheselli, Università Ca' Foscari Venezia

Keywords: Sampling, data collection, data journey, epistemology, methodologies

Sampling and data collection practices are central to how technoscience understands and intervenes in the world, yet these are often overlooked aspects in STS scholarship. Sampling practices embody particular assumptions about what counts as valid knowledge, whose perspectives matter, and what constitutes good science. These practices involve key decisions about who or what to include in a study and how to gather that data, often influencing the reliability and validity of scientific outcomes. These decisions are not merely technical—they are shaped by broader historical, cultural, and ethical contexts. The design of sampling methods can impact both the research itself and the communities involved, raising questions about justice, representation, and equity. How do sampling and data collection practices contribute to or complicate the making of "good" technoscience? What ethical, epistemological, and practical challenges arise from how sampling is conducted, and how do these methods shape knowledge production and its broader social impacts?

We welcome contributions that engage with sampling/data collection practices in a broad range of scientific fields and contexts, to discuss the origins of data in scientific practice and the relevance of critically engaging with these practices for the possibility of "good" technoscience. We encourage contributions that deal with themes such as:

- how sampling practices reflect and shape valuation of "good" science;
- the relationship between sampling methods and scientific outcomes;
- ways that different communities experience and engage with data collection;
- the role of values and ethics in sampling decisions;
- tensions between standardization and local knowledge in data gathering;
- cross-cultural perspectives on data collection;
- the complexities of medical sample collection, from the intricacies of clinical trials to
- the ethical considerations; surrounding human specimen collection;
- the porous boundaries of the lab;
- sampling as embodied practice;
- the mutual shaping of sensory experience, knowledge making practices and technology in use;
- interdisciplinary collaborations built around fieldwork and sample collection processes;
- sample collection and citizen science;
- the role of the STS researcher in sampling practices;
- metaphors, narratives and discourses of sampling practices;
- how sampling practices shape understandings of ecosystems and environments;
- local sampling and planetary thinking;
- sampling, extractivist economies and natural resources exploitation.



11 JUNE 2025 09.00 - 11.00

ID 172 - Dialectical Primatologists and the Negotiation of Liminality in Peri-Urban Ecologies

Kymerley Chu, Princeton University

Keywords: Conservation, Decolonial STS, Multispecies Theory, Urban Ecology

How do plantations and their extractive-based economies impact the ways scientists, agricultural workers, and urban residents interact with animals? How do conservationist biologists generate primate behavioural data as they study free-ranging monkeys in fragmented landscapes experiencing urban expansion and monocrop cultivation? What kinds of human-monkey interfaces are being legitimised or obscured in citizen science fieldwork?

In Malaysia, natural scientists must navigate the government's research permit acquisition system, where all animals are considered legal subjects of governance. In exchange for their conditional research permits, scientists precariously situate their primate behavioural data to government officials as a series of normative, moral, and political contributions in aiding Malaysia's plantation and urban development activities. This sampling data collection process reshapes government, scientific, and agricultural activities as incubators for novel forms of human-monkey sociality including the simultaneous portrayal of monkeys as 'ecologically destructive pests' in development activities, ecological knowledge co-producers, and as displaced families. Scientists must contend with simultaneously framing free-ranging monkeys as agential subjects and as behavioural data objects during fieldwork.

In landscapes featuring peri-urban (land featuring combined urban and rural fringes) ecologies, local scientists, through the approach of citizen science, conduct behavioural research methodologies such as focal sampling and scan sampling. Focal sampling situates individual monkey subjects as distinct leaders, while scan sampling records the collective behaviours of primate social groups over regularly timed intervals. When disseminating sampling data to the public, local scientists interpret specific behavioural observations as a proposed continuum of conservationist solutions in human-monkey interfaces usually framed as inherently and innately antagonistic. Overall, scientists, agricultural workers and urban residents perceive monkeys as co-inhabiting multiple relational identities between various communities despite living in the same ecosystems and cultural landscapes.

In conclusion, this paper is a broader call for social and natural scientists to reconceptualize agential non-human beings as co-inhabiting a relational series of social identities unevenly deployed across dynamic landscapes, shifting local environmental stressors, and in ecological knowledge.

11 JUNE 2025 09.00 - 11.00

ID 354 - The Emperor's New Crowds: Unveiling the Strategic Uses of Collective Wisdom

Niccolo' Tempini, University of Exeter

Laura Savolainen, Helsingin yliopisto

Keywords: Crowdsourcing, sampling, judgement, formalisation For a long time, crowdsourcing platforms have been benefiting from the appeal of exploiting the "wisdom of the crowds" (WoC) to generate valuable data: WoC is potential to any collection of diverse individual observations. This assumption has taken such deep seat in contemporary discourse of web-based knowledge practices, that it goes often unquestioned by scientists and users of such platforms. It is also practically as well as ideologically implicated in the development of groundbreaking AI innovations such as LLMs.

In this paper, we explore the wisdom of the crowds assumption. We ground our analysis in the empirical data from two research projects conducted by each of us individually. In one, we interviewed scientists who use crowdsourcing platforms to study how they think about their work with the crowd and the data thus



generated. In the other, we observed developers of a participatory research platform to study how they shaped website design and workflows to source high value data from the community of users.

Crowdsourcing practitioners do not talk about WoC as an assumption. Rather, it is treated as an informal law of nature, a phenomenon that is regularly observed, when some basic conditions are satisfied as to the nature of the crowd, and which can be exploited to one party's benefit. However, WoC is anything but an informal law of nature.

The most common articulation of the WoC assumption is, roughly, the claim that a high number of observations from non-experts (average) humans, will average towards the accurate. In its stronger forms, the WoC is stated as superior expert judgement. It assumes that the aggregate observation of a crowd of non-experts will be, equal if not more, accurate than the judgement of an individual professional expert.

However, comparisons cannot be straightforward and when put to the test, WoC is revealed as a convenient fiction. There is great heterogeneity in the way researchers relate to the crowd and the tasks that it is asked to complete. Researchers routinely transform the problem they are trying to address to fit the mould of a crowd-sourcing template of data generation. They can assess, dismiss, and overrule crowd work. This is needed not the least because of a great amount of sampling uncertainty with respect to the crowd which undermines confidence in a straightforward comparison of crowd wisdom vs expert judgement or in an explanation as to the causes of observed crowd (in)accuracy. WoC is, at best, a misplaced interpretation of the causes of possibly accurate crowd outputs.

How is an influential interpretation of crowdsourcing practices such as WoC can be unwarranted yet continue to be pursued? The answer to this question is that those practices are not concerned with accuracy with respect to an external standard, but rather, they produce the standard that they are meeting. While a better and more rigorous standard is most often conceivable and desirable, practical constraints mean that the adoption of a standard that is endogenous to the WoC methodology is going to be enough for common applications.

11 JUNE 2025 09.00 - 11.00

ID 380 - Is digitisation of Ecology actually "good"? Digital technologies' role in sampling and laboratory practices, and their epistemic implications

Stefano Oricchio, Università di Napoli Federico II

Giuseppina Pellegrino, Università della Calabria

Keywords: digital technologies, ecology, sampling, bio-monitoring, ecology of infrastructures

Digital information and communication technologies (ICTs) have long since become a constitutive part of science, being deeply intertwined in every step of the research process. Ecology - an interdisciplinary field, acquiring an ever-increasing relevance in times of climate change - is no exception to the digital transformation of science. Despite this, the implications of digitisation of ecological knowledge are still neglected and little discussed.

Drawing on semi-structured interviews to academic ecologists such as entomologists, hydrologists, botanists, and ecotoxicologists, this contribution adopts an STS theoretical framework, namely the Ecology of Infrastructures approach (Star and Ruhleder, 1996) - and firstly aims at illustrating the role of digital ICTs in the ecologists' fieldwork and laboratory practices. Even if digital technologies constitute an invisible and mundane research infrastructure, the interviews showed how every step of data collection, management, analysis, storage and dissemination is deeply digital. This primarily applies to laboratory practices but is consistently relevant for fieldwork and sampling techniques, which are densely manual and sociomaterial as much as supported or enhanced by different kinds of digital ICTs (e.g. simple cameras, automated data loggers, remote sensing devices). The assemblage narrated reveals tensions and ambivalences: digital technologies amplifies both precision and ambiguity, as researchers navigate challenges in devices usage, data contextualisation, and infrastructure access, in line with memory and forgetting practices in the



sciences (Bowker, 2008). This provides a first argument on whether digitizing ecology is actually a "good" practice.

Secondly, the contribution further delves into the epistemological implications of digital ICTs on the evolution of Ecology. The "informatisation" of the environment, i.e. the framing of every ecological process in terms of informational flows that can be understood and described by scientists (Bellamy Foster and Clark, 2008) - is a process widely acknowledged by the science history literature, but a critical discussion of its epistemological implications is not yet fully developed. From mathematical models to the widespread reference to species as bio-indicators, the digital evolution of Ecology suggests an increasing tendency to evaluate - rather than describing, as once it was the case - the environment. Such an epistemic shift leads to a potentially controversial and paradoxical politicisation and normativisation of Ecology. This can result in a rejection of scientific knowledge and, on the very political side, a kind of paralysis that leads government and society to scepticism and inaction, as evident in the case of global climate change controversy (Edwards, 2010). This provides a second argument on the supposedly "good" aspect of digital ecology, in epistemological and ethical terms.

References:

Bellamy Foster, J., Clark, B. (2008) The sociology of ecology: ecological organicism versus ecosystem ecology in the social construction of ecological science, 1926-1935, in *Organisation & Environment*, 21, 3: 311-352.

Bowker, G.C. (2008) *Memory Practices in the Sciences*, Cambridge, MIT Press.

Edwards, P. N. (2010) *A vast machine: computer models, climate data, and the politics of global warming*, Cambridge, MA, MIT Press.

Star, S. L., Ruhleder, K. (1996) Steps toward an ecology of infrastructure: design and access for large information spaces, in *Information systems research*, 7, 1: 111-134."

11 JUNE 2025 09.00 - 11.00

ID 453 - The production of sound evidence: cosmologies and epistemic virtues in sampling, clustering and categorizing in a precision oncology project

Fabio Gasparini, CNR

Lorenzo Beltrame, Università degli Studi di Trento

Keywords: sampling, clustering, epistemic virtues

It's a critical moment in the laboratory of Dr. Nora Rubicone. She and her research group have to select a sample from the cohort of patients undergoing a prostate cancer therapies. Patients have to be selected, clustered and assigned to three groups based on their response to treatments: short, medium, and long responders. Although each patient represents a source of data, the patients in themselves do not embody any form of order. There is no evidence in them.

Evidence is not inherent in data, nor does it emerge automatically from them: evidence must be constructed by cleaning data, which involves selecting a sample of patients and clustering them into categories that can enable robust, solid and statistically significant analysis. According to Boumans and Leonelli (2020), cleaning by clustering is a form of epistemic ordering - a practice of imposing order and intelligibility that turns the disorder of raw data into the basis for the production of sound evidence. The organisation of data through sampling, clustering, and categorisation does not reflect some clear structure present in the external world and already inherent to the cohort of patients. This structure results from a cosmology reflecting the social organisation of those producing the order.

In this paper, through the ethnographic analysis of the processes of sampling, clustering, and categorisation of responders in a translational oncology laboratory, we show how the construction of a sample reflects an active tension toward the natural order (cosmology), which is in a homeostatic relationship with the social order of the actors engaged in data-cleaning work. While following the insights of Boumans



and Leonelli (2020), this work demonstrates that the process of ordering through sampling, clustering, and categorisation does not exclusively follow "specific research questions and goals" but is influenced by "epistemic virtues" (Daston and Galison 2007), internalised by the actors. Following the actors in their effort to decide how to sample, to cluster and to categorize patients, we will show how these epistemic virtues (that "appeal to ethical values, as well as to pragmatic efficacy in securing knowledge" Daston and Galison 2007, 40) shape not only actors' data practices, but also their understanding of robustness and statistical significance.

11 JUNE 2025 09.00 - 11.00

ID 655 - Seeing like a water fountain: reimagining data collection in London's #OneLess refill pilot

Nicole Vitellone, University of Liverpool

Keywords: Environmental experiments, water, ethnomethodological research techniques, everyday practice

This paper engages the capacity of data collection practices in my research on London's #OneLess refill water fountain pilot as central to understanding how technoscience intervenes in the world. Drawing on qualitative interviews with refill water fountain users engaged in the #OneLess pilot, the paper examines the data gathering devices used by ethnomethodologists for studying everyday material practices and settings. Comparing the system change perspective of practitioners responsible for the design and evaluation of the #OneLess campaign to reduce single use plastic water, with the ordinary practical methods of research participants engaged in the use of the refill water fountain pilot in everyday settings, the paper addresses the contribution ethnomethodological data collection practices make to STS research and knowledge production. In so doing we tease out the distinctive contribution of ethnomethodological research techniques to data gathering with human and nonhuman actors and what they contribute to the process to realise environmental infrastructural transformation. By studying everyday mundane practices as an apparatus to render technoscientific phenomena explorable and analysable in local experimental settings, the paper outlines what the interview technique and detailed transcription contribute to data collection and the methodological challenge of engaging practitioners from different professional groups and across disciplinary boundaries in reimagining what counts as valid knowledge.

11 JUNE 2025 09.00 - 11.00

ID 718 - Citizen science in Chile: from highlighting socio-environmental challenges to reflecting on social impacts

Lucie Le Goff, Pontificia Universidad Católica de Chile

Keywords: citizen science, co-production, sampling practices, social impact, Chile

Citizen science covers a range of projects that include local populations in the production of knowledge. Their involvement is gradual, ranging from contributory, collaborative and co-created projects to those that are independent of academic scientific institutions.

The aim of this presentation is to offer a descriptive analysis of Chilean citizen science, which has developed over the last ten years. The results presented are based on an analysis of a historical register of citizen or community science projects (150) created in Chile to date. This work was carried out as part of the Millennium Nucleus working group on Citizen Technoscience for Socioenvironmental Transformation (CITEC) (<https://mileniocitec.cl/>).

We want to examine the role of collective data sampling and citizen participation in the production of 'good' science. What contribution does citizen science make to the production of good science? Does the participation of the populations concerned make this type of science more valid and ethical or, on the



contrary, does the way in which the data is produced make it more open to criticism? Does the knowledge co-produced have a greater potential impact on resolving local socio-environmental challenges? Does the sample collection of citizen science make it possible to produce useful and robust data for guiding public policies?

The data collected by these projects produces knowledge that is rooted in the day-to-day reality of local areas, and highlights issues of interest to local residents and/or scientists (animal and plant biodiversity, terrestrial and coastal ecosystems, quality of the environment, risks, adaptation to climate change, etc.). They therefore offer an approach to society's aspirations regarding relations between humans and non-humans. There is a diversity of disciplines (natural, human and social sciences) and data sampling practices involved (environmental monitoring, geo-referenced photographic register, knowledge dialogue, etc.). As most of the projects focus on biodiversity, there is a marked predominance of online data collection platforms via mobile applications, which collect quantitative databases. The use of digital technologies allows for a wider sampling scale, with numerous regional, multi-regional and national projects. At the same time, the representativeness of contributory projects raises questions about the real social impact versus the utilitarian drift of involving local residents. However, it is evident that some projects take the form of epistemic justice, whether they advocate recognition of indigenous knowledge, or whether they offer populations exposed to risks the means to measure the quality of their environment.

There is no standardisation of data collection for the same subject of study (e.g. water quality). Their duration is often short, in line with their objectives and above all with their funding. However, the sustained production of data over time is the key to achieving a significant social and environmental impact. However, these data sets offer the advantage of producing knowledge in areas where there is no public data, and to provide discussion tools for local decision making.

It is in the interests of Chilean citizen science to capitalise on and share learnings in order to contribute more effectively and sustainably to a fairer socio-environmental transition.

