Workshop "Data – Difference – Diversity. Technologies of Differentiation in the Life Sciences", MPIWG Berlin (IRG-III), 24-26 November 2011,

organized by: Susanne Bauer, MPIWG, Christine Hanke, KHM Cologne,
Sabine Höhler, KTH Stockholm

Abstracts:

Thursday, Nov 24, 14:30 – 18:00

Session: Data – Difference – Diversity: Perspectives

Thick Data, Time, and the Politics of Abduction

Michelle Murphy, University of Toronto

This paper engages the layered and thick data collected on cholera over the twentieth century from a single field site in Bangladesh. This rural field site is not only the most important research location for the study of cholera, it is also home to world's oldest demographic surveillance project and a crucial location for research on family planning. The paper develops the notion of abduction – both in its pragmatist sense but also in its affective and coercive meaning, to think critically about looping practices that repurpose data towards changing conjectural futures. The paper critically entangles contemporary ecological and structural studies of cholera, as well as the contemporary practices of historians of science, with the politics of abduction.   The paper reflects on the postcolonial political economy of infrastructure that generations of research and data accumulation has built at this site. In conclusion, the paper argues for abduction as a felt and temporal condition of data politics that replaces a sense of historical time, and which posits the past as not that which has happened but which we can know and reoriented technically towards non-innocent futures.

Differentiating Infrastructure

Brit Ross Winthereik, IT University of Copenhagen

Information infrastructures are crucial to the operations of modern society and its organizations. They are also tricky analytical objects. Historian of technology, Paul Edwards, reminds us that infrastructure was originally a military term designating "fixed facilities". Infrastructure was seen as the underlying basis on top of which or through which a society or an organization operated; it covered such areas road nets, waterways, power grids and even public institutions such as schools and prisons (Edwards, 2003: 187: quoting the American Heritage Dictionary).
This paper introduces an approach to information infrastructures that sees them as 'facilities' whose boundaries are at one and the same time fluid and fixed.

Based on fieldwork in an environmental NGO that focused on monitoring and evaluation technologies and their role in developing and maintaining collaborative partnership with distant actors, the paper attends to information infrastructures as both the underlying basis of global collaboration and as weedy compositions that only occasionally allows for collaboration and partnership. The paper proposes to differentiate the notion of infrastructure by installing a double vision that allows for keeping different imaginaries of infrastructures in the same picture. Anna Tsing's notion of the "swidden" is useful in this regard (Tsing, 2005). "For an observer used to imagining agriculture as cleanly weeded lines of corn, wheat, or tomatoes, but not all tossed together", Anna Tsing writes, "the most amazing thing about a Meratus swidden field is the extraordinary number of plants growing together in the same small spot" (2005, 165). This scene of enormous variety is likely to be interpreted by "visitors from the city" as "'natural' forest" (174), but it is not. Indeed, most of the variety represents cultivated variety not wild nature, yet it is seen as wild. As an infrastructured space, whose infrastructure is nevertheless hard to identify for outsiders, the swidden urges us to once more consider how to write about what Leigh Star has also alerted us to: "one's persons infrastructure is another's brick wall". Contemplating the possibility of narrating infrastructures as both cultivated and wild implies considering how to write a 'switch' in one's ethnographic accounts of infrastructures that allows for seeing both the orderly and the weedy (cf. Winthereik & Verran, forthcoming).

Measurement, Data Analysis and Visualisation – Remarks on Medial Performativity

Christine Hanke, Academy of Media Arts, Cologne

Measurement, data storage, accumulation, and evaluation, statistics and visualisation are key technologies of recent life sciences. My paper will focus on the media specific effects and performativities of these technologies in the procedures of generating difference and diversity. From a media theory perspective, the productivity of measuring and visualizing approaches is linked to the epistemic logic of two 'basic media': number and image. Starting with a characterisation of the number and image medium, I will sketch out effects bound to these media in measurement, databases, statistics and data visualisation. I will discuss the epistemic function and the power effects of these different medial approaches: as they unfold specific and sometimes unforeseen and indeterminate effects within the constitution of differences and diversity: It is the very medial performativity that lets us 'see' differences in a specific way, but involves also indeterminacies and blurrings of boundaries and categories. Taking the medial performativity in measurement, data analysis and visualisation into account I will propose an approach towards diversity that is reflexive of the involved medial technologies.

Protean proteins: methods in pursuit of identity and difference

Adrian Mackenzie and Ruth McNally, Lancaster University

Things become more real when we forget how they are made. But what happens to making-real through-forgetting when the very methods of knowing and making shift? How we know something today may not be the way we know it tomorrow. Contemporary biology is instructive in this regard. It has turned to high-throughput, data-intensive instruments for prising apart living things. This turn has seen laboratory life undergo radical transformations in career structure, infrastructure and global flows of data. Yet these changes bring such a tumult of different ways of knowing and making that it can sometimes be hard to comprehend what has been discovered. Signs of this difficulty can be seen in the visual devices which biologists use to show the results of their own knowing and making. We discuss a visual digital device called the 'cluster heatmap,' and focus on one atypical use of it in the field of proteomics. Rather than displaying experimental data, in the case we describe a heatmap was used to explore the differences and similarities between experiments focused on the proteins in human blood plasma. On the one hand, this heatmap could be read as an acknowledgment of the performativity of scientific knowing. It might be saying: On the other hand, this acknowledgement is itself performed by recourse to a panoply of methods, metaphors and data modelling techniques drawn from various disciplines. The findings presented in the heatmaps themselves veer between an insistence on the need to find pattern in data using all means available, and an almost inadvertant admission that much depends on who made that data and where. The continuous invention of methods destabilises the real, the constant destabilisation of methods re-invents the real.

Friday, Nov 25, 9:30 – 17:30

Session: Data Routines I: Stocks and Flows

The Data Deluge: Reinventing Authorship in the Data-Driven Sciences

Bruno J. Strasser, Yale University

From the pages of *Wired* to those of *The Economist*, *Nature* and *Science,* commentators have characterized the early 21st century as a moment in history defined by a "data deluge". This unprecedented amount of information, they argue, will deeply transform the scientific enterprise, turning it into a "data-driven" practice. This paper is an attempt at bringing these claims into historical perspective. It argues that our current grand narratives about the history of the life sciences are inadequate to make sense of current developments. The current "data deluge" is not unprecedented and thus the past can offer clues to understand the present. As Robert Darnton has convincingly argued, "every age was an age of information, each in its own way". This paper will first survey

some current claims about "data-driven" science, then it will discuss how the current historiography obscures, rather than illuminates, this development and propose a new framework, centered in the historical role of "collecting sciences". It will then show that one of the essential tensions in the "collecting sciences" (including the current data-driven sciences) has been the negotiation of individual credit and collective participation. Taking examples from crystallographic and genomic databases, it shows how these tensions were resolved by redefining the very meaning of scientific authorship.

Search channels, background isolations and statistical significance: Ways of detecting the Higgs particle in LHC data

Martina Merz, University of Luzern

The Large Hadron Collider (LHC) has taken up operation at CERN, the European Laboratory for Particle Physics, in summer 2010 after a preparatory and construction phase of almost two decades. Since then the discovery of the Higgs particle seems imminent. First candidate events have been presented at a scientific meeting in July 2011 but up to date (late August 2011), CERN physicists have not presented conclusive evidence for the existence of the Higgs particle. What is at stake?

The Higgs particle is the last missing building block in the well-established Standard Model. Its discovery is expected to inform a more general theory of particle interactions including gravitation. However, detecting the Higgs presents an enormous challenge. Not only does its presumed mass range require the construction of ever more powerful accelerators such as the LHC. Also the work of data analysis has to overcome major obstacles; how physicists meet this challenge will be the focus of my presentation.

The LHC produces a data flux of several petabytes (1015 bytes) per year among which the exceedingly rare (hypothesized) Higgs events have to be uncovered. "Discovery" relies on statistical reasoning facing the challenge that statistical fluctuations have to be distinguished from the sought-for signal. The Higgs particle is produced in the collision process by a variety of physical processes and, once produced, decays into various "channels." In the course of data analysis, physicists attempt to isolate data associated with these decay channels from the dominating background processes.

Epistemic strategies of physics analysis (such as the choice of search channels or issues of statistical significance) are socially embedded in various policies, which organize this work as a multiply distributed work among thousands of people and major physics laboratories. These policies aim at producing credibility at different interfering levels (of individuals, of software, of results on other channels, between collaborations, etc.) and at getting the work done. In my presentation, I will focus on the different strategies of credibility production and the associated policies with the aim to better understand the multi-layered practice of contemporary particle physics as being at the same time data- and model-driven.

Respect for Epistemic Diversity: A Guiding Principle for Biological Cyberinfrastructures

Sabina Leonelli, University of Exeter

Computational tools developed to disseminate scientific data over the internet, often referred to as 'cyberinfrastructures' or 'e-science tools', are widely seen as a solution to the increasingly globalised and distributed nature of research networks. E-science tools help researchers to communicate and co-operate across geographical, disciplinary and cultural boundaries, thus facilitating the creation and development of large networks able to overcome the physical fragmentation of research in laboratories spread around the world. Indeed, the growth of e-science is often portrayed as intertwined with the growth of 'big science'. This paper probes the relation between e-science and big science by examining two specific e-science tools used to serve model organism research and cancer research: The Arabidopsis Information Resource (TAIR) and the Cancer Biomedical Informatics Grid (caBIG). In both cases, the science in question is 'big' in several respects: the number of people and disciplines involved, their geographical spread, the economic and scientific resources utilized and the expectations piled upon research results. I argue in order to be useful to such large and complex communities, databases such as TAIR and caBIG need to cater primarily for local, rather than global, needs. In other words, rather than trying to establish standards that uniformly apply across all user disciplines and communities, what the curators of these resource struggle with is the need to preserve and encourage the use of local standards, terminologies and practices that mirror the diversity of the epistemic cultures of their users. Respect for epistemic diversity is thus a crucial guiding principle for the design and development of cyberinfrastructure within the biological and biomedical sciences. This paper explores the contradictions and difficulties underlying such an ambitious project.

Chains of inference in infectious disease transmission models: What happens to diversity?

Erika Mansnerus, London School of Economics

Increased use of modelling methods in infectious disease epidemiology seems to provide an easy way to produce predictions of pandemic outbreaks or estimates of transmission rates. While models and simulations gain more and more interest from sociological and philosophical communities, hardly any attention has been paid on the effects of model-based evidence and its use in shaping our understanding of disease transmission. It is often said that models should remain as simple as possible. This methodological efficiency threatens diversity. By simplifying the modelled phenomena, information is lost. This, however, is

not the only point during the modelling process that has an effect on diversity. This article analyses the chains of inference that organise knowledge of disease transmission in modelling and discusses what happens to diversity. The article is based on case studies on infectious disease modelling.

The starting point is a set of simple transmission models that applied homogeneous mixing within population. In other words, these models represented individuals as 'gas molecules' mixing in a test tube. Obviously, this simplified form helped to solve the mathematical problem of calculating transmission through contacts but at the cost of more accurate understanding of how individuals make contact and transmit infections. Having established simplified understanding of disease transmission, more elaborate dynamic transmission models were developed to address the shortcomings. These simulation-based tools acknowledge the heterogeneity of the population through age and contact structures that are embedded in the model. They seem to be capable of accommodating diversity. The more sophisticated the model, the more detailed understanding of the transmission dynamics is gained. This accuracy with the model-based representation of the infection and its spread in a population, however, is compromised when model-based evidence is communicated to the decision-making processes. The nuances of model assumptions and the potential problems of datasets used for calibrating the model are lost when the results are briefly introduced outside the modelling community. This change from a detailed model-based representation of transmission to its simplified, graphical or animated form affects diversity. In this stage, however, diversity is not only compromised because of technical or computational restrictions. The elements in the models (such as 'missing data', and uncertainty) that are not known outside the modelling community and that are difficult to communicate play a key role here. Through these examples, the article shows how diversity is represented in the chains of inference that take place in models and discusses and what are the consequences of the fluctuation between simplification information and increasing heterogeneity in models.

Session: Tracing/Producing Difference

Taste Sensitivity to Phenylthiocarbamid Around the Globe: Collecting Data on a New Genetic Marker in the mid-20th century

Veronika Lipphardt, Max Planck Institute for the History of Science, Berlin

The ability to taste the substance Phenylthiocarbamid (PTC) is one of the best-studied genetic characters in humans, second only to the ABO blood group system, for its inheritance follows a strict Mendelian pattern. Humans can or cannot taste PTC as bitter and will give yes/no answers, with no intermediary conditions. Compared to blood groups, PTC taste sensitivity is relatively easy to assess.

Discovered by chance in 1931, PTC taste sensitivity soon became a focus of genetic transmission studies as well as of genetic-anthropological studies.

Scientists had soon found out that the percentage of tasters differed between populations; thus, this new marker seemed appropriate for characterising not only individuals, but also populations. Students of population genetics and human evolution set out to gather data on PTC sensitivity from populations around the globe, intending to compare their results and to find clues on evolutionary developments. Particularly in non-Western countries, studies were often carried out alongside medical studies (although PTC taste sensitivity is considered non-pathological) or as part of already established anthropological research endeavours. Two decades later, a considerable mass of data had been compiled.

However, the population data did not reveal clear patterns. Instead of questioning the stability of population categories, scientists suspected that not enough data had yet been collected and that the technical procedures needed to be standardized. Furthermore, the collected data revealed characteristics of the marker itself that rendered it more unstable and complex: It varied with sex; in some populations tasters would taste it with the same clarity, but only in much higher concentrations than in other populations. These findings triggered new studies and the collection of ever more data for the purpose of the marker's characterization.

This paper aims to trace the history of the marker's discovery, of its calibration and stabilization in the light of population mass data. Furthermore, it discusses how new markers could be integrated into existing data collections and data-collecting endeavours, and how new research questions were generated from the data. The paper thus addresses the interplay between markers, populations and data, viewing each of them as epistemic object and epistemic instrument.

Tracing differences: human chromosome research, 1960s to 1970s

Soraya de Chadarevian, University of California, Los Angeles

The standardization of the human karyotype in the 1960s not only distinguished every single chromosome in a chromosome set. It also and most importantly provided the basis for reliably describing differences between karyotypes. It was differences that mattered. Differences in the number or structures of chromosomes could point to variations between population groups, to 'normal' genetic variation, to disease patterns, to the mutational effects of radiation or other environmental agents. Large screening programs and ever more elaborate technologies were set in place to capture these differences and to gain clues about the mechanisms that governed them. In my paper I will discuss some of these projects, focusing on the material and visual practices to record differences, on early attempts to computerize the process to deal with large data sets and to the institutionalization of centralized tools to collect and store the information.

Suspect families: DNA kinship testing and its implications for the concept of family

Thomas Lemke, Goethe-University, Frankfurt

Since the 1990s, many countries around the world have begun to use DNA testing to establish biological relatedness in family reunification cases. Family reunification refers to the right of family members living abroad to join relatives who hold long-term residence permits in a given country. While this right has been an integral part of many countries' immigration policies, the current trend among host countries seems to favour more restrictive family reunification policies. Even if applicants possess the required documents, immigration authorities often reject the information as they question the authenticity of the documents. In this context, many countries resort to DNA tests to resolve cases in which they consider the information presented on family relations to be incomplete or unsatisfactory. Today, thirteen European nations have incorporated the use of DNA testing into decision-making on immigration. Using Germany as exemplary case study, I discuss the social and ethical aspects of DNA testing for family reunification and their implications on the concepts of family. DNA kinship tests for immigration purposes establish and strengthen a nuclear family model. Family is increasingly understood as a biological relation while social dimensions such as affections and the aspect of caring for each other are devaluated. This biological idea of family contradicts the more pluralistic concepts of family in western European countries, e.g. same sex partnerships or families with adopted children. The argument is based on interviews with representatives of NGOs and immigration authorities, lawyers specialising in family and immigration law, geneticists working in DNA laboratories and those applying for family reunification.

Biodiversification: Service Ecology's Economy

Sabine Höhler, KTH Stockholm

In the second half of the twentieth century concepts of diversity not only circulated in the life sciences and humanities but they also linked ecological and economical thought. This paper explores the meanings and functions of biodiversity in what I call the rise of a global 'service ecology.' I employ this term to point to a novel perception of nature as service provider, supplying ecosystem goods and services with the primary purpose to support human life on earth in a sustainable fashion. Service ecology, I argue, materialized in the 1970s along with the servitization of products and supplies that characterized the service economy, the growing tertiary sector of national and global economies. With the servitization of nature biodiversity turned into a key factor for the regulation and provisioning services of ecosystems. As part of nature's "life support

functions" biodiversity denoted the complex and interdependent ecosystem infrastructures to buffer and dynamically adapt to external pressures by sustaining species richness, genetic wealth, and ecosystem diversity. The invention of biodiversity in the 1980s and its institutionalization with the UN Convention on Biological Diversity in 1992 effectively connected diversity to (human) survival; moreover, diversity thinking supported the global ecological and economic perspective of a "resilient" nature that mirrored the late modern change towards the diverse and flexible achieving society. Similar to the human-capital-based service economy, biodiversity became the central dispositive of a service ecology built on natural capital. In the paper I will study the recent ecological imperative of managing nature's "portfolio" in terms and practices of 'biodiversification': conserving assets and banking value, creating and realizing benefits, assessing and distributing risk, optimizing performance and stability, and assuring response diversity and stress resistance of ecosystems. More generally I aim to discuss how the discourse on diversity informed and expressed the argument that sustainability needs markets operating on the valuations and transactions of natural capital stocks.

Saturday, Nov 26, 10:00 – 16:00

Session: Effects and Genealogies

New Adjustments between Theory and Empirical Data

Gabriele Gramelsberger, FU Berlin

Computer based simulation is not only a method for generating 'in-silico' data, it is also a method for operating on data–usually received from measurement. These measurement data import empirical data into simulations, which are seen as experiments on theory. These data allow to adjust simulation models with reality as well as to empirically evaluate the simulation results. However, this differentiation between theory and empirical data is challenged in two ways: Firstly, because in-silico data are theoretical as well as empirical, they generate a new and hybrid form of scientific data. Secondly, because today's practice of data detection increasingly turns empirical data detection into a simulation based data generation. Of course measurement data have always been theory-laden but the introduction of the computer not only enabled computer based simulation as a new method, it also increasingly equipped measurement devices with computer chips respectively machine algorithms–thus increasingly subordinating measurement to theory. These algorithms generate new data by generating new differences on a higher level. Such strategies are called data retrieval in science, e.g. for re-interpreting the raw data again and again. Thus, a limited data basis can lead to infinite new data sets. In meteorology data retrievals are used for satellite data as well as for the large repositories of historical data. In biology and chemistry data retrievals are used, vice versa, to re-interpret theoretical data for synthesizing new proteins or molecules. The paper will present a philosophical perspective on these new data generating

practices by posing the question, whether new forms of data generation dissolve the traditional adjustment between theory and empirical data and establish a new line between both.

Genealogies of the Interface: From Representing Information to Data Visualization in Cybernetic Thought

Orit Halpern, New School for Social Research, New York

This paper examines how ideas of communication and cybernetics transformed concepts of measurement, data, and representation in the cognitive, computational, and social sciences after the Second World War. Linking together a number of cybernetically informed individuals who interlocuted with one another—Norbert Wiener in cybernetics, Warren McCulloch in neuro-science and neural nets, Kevin Lynch and Gyorgy Kepes in design and urban planning, Karl Deutsch in political science, and George Miller in cognitive science--I will produce a speculative genealogy of our contemporary attitudes to data visualization, surveillance, space, and the interface.
In these archives we can find traces of a broader epistemology that made measurement about behavior, interactions, and scaleability and reconfigured the relationship between representation and analysis as a single and simultaneous process. Replacing older ideals of taxonomy, ontology, and archiving, these practices produced novel objects of study, and new sites for intervention, such as the environment and cognition by way of measure. The idea of data shifted from being a representation of an external world, to being a tool for producing a new world, just as analyzing, collecting, and displaying data became part of the same process in a manner anticipating our contemporary notions of "visualization".
The result of these changes was that older relationships between normativity, population, and territory were reformulated and replaced by the terms of capacity, network, and process with an impact on everything from psychology to urban planning. In the work of these figures, therefore, lies a history of our contemporary faith that more information is better, interactivity equates with thought, and that the constant gathering of data is a virtue. This is a suggestive genealogy, then, of how we came to assume data is relational, ubiquitous, performative, and interactive.

Biosocial classifications and the emergence and visibility of "new societies"

Felix Keller, University of St. Gallen

"New societies form along newly recognized (or, at any rate, newly asserted) biological or genetic lines, forging new alliances and loyalties", wrote Ian Hacking some years ago. But what might the evidence be of these new societies, what "stuff" are they made of? Obviously, the "genetic imperative" has gained reality in medicine and kinship identification. Moreover, there is some kind of a rebound of a biological understanding of society, promising soon to identify genetic markers determining social realities such as criminal behavior or sexual preferences. But this kind of information, if it can ever be reliable at all, contextualizes societies and individuals as they are, delivering information about their biological background and, perhaps, limiting the trust in cultural forces. There would seem to be no reason to call them "new societies". Kinship, for example, has always been a mix of biological and social information.
Is there any argument for introducing a new concept of societies, in view of the use of biosocial classifications? Beside the endless debates on the degree to which the biological determines the social, one can also observe a shift from the "bio-deterministic" to a more probabilistic way of thinking about the biological and the social. In this thinking, biological and social classification are not linked in a casual manner, but conceptualized as interfering spaces of probabilities *beside* the concrete individual or social group. The most significant shift in this direction can be seen in the attempt to create *biosocial surveys*. The intention of biosocial surveys is to link databases of classical social surveys with databases consisting of biological information in a probabilistic manner. The purpose of biosocial surveys is obviously not a biological ontology of social or individual living, but a new tool to simulate and control risks, to manage biological and social diversity in a new flexible "biosocial space". The importance of surveys as tools to represent and manipulate modern societies is widely known and shown. Looking into the concepts of this new kind of survey, contextualizing it with the history of statistical thinking and biosocial classification, this paper explores whether these new biosocial survey strategies can indeed be seen as indicating the emergence of "new societies" that individual beings have to live with.

Making Time: On Race and Relations in Cows and Humans

Amade M'charek, University of Amsterdam

Time is central in studies of diversity. The so-called molecular clock is a crucial device to estimate evolutionary changes in a species or in the DNA. This clock has to be 'calibrated' in order to interpret the mutation rate and to relate 'mutation events' to a historic time. In this sense the molecular clock is a crucial means for understanding genetic diversity.
In this paper I want to explore time as a technology of making similarities and

differences. However, contrary to the evolutionary approach sketched here, which relies on a model in which time flows in a linear mode,[1] I will take my inspirations from Michel Serres's notion of topological time, a folded and crumpled time. In Serres' universe time does not flow, but percolates. It does so in a rather turbulent, chaotic and anachronistic way. Serres gives us a stunning example of anachronism by recounting a story of two brothers, in their seventies, who were burying their father aged some thirty years. When these two brothers were still young boys, their father had an accident in the high mountains and went missing. Over the decades to come his body and youth were conserved in the coldness of the glacier valley.[2]

This example is somewhat exceptional, but the mechanism and the folding of time is much more common as I will argue. I will take some examples from two cases, human genetics and archaeology on the one hand and cattle breeding on the other, to examine how time is deviced to make relations including racial ones.

[1] On the naturalization of linear time, see e.g. Geoffrey Bowker **(**2006) *Memory Practices in the Sciences* Cambridge: MIT Press; Zara Mirmalek (2009) "Working Time on Mars" *KronoScope*, Vol. 8, No. 2., pp. 159-178

[2] In Michel Serres & Bruno Latour (1995) *Conversations on Science, Culture and Time* Ann Arbor: University of Michigan Press.