"Extended Report" of the 1st Workshop of the BCSSS on the *Epistemology of BD* and Perspective of the 2nd Workshop in Spring 2019

WHAT DOES BIG DATA TELL US? - Epistemological questions of systems science

Author: Felix Tretter (BCSSS), Bertalanffy Center for the Study of Systems Science (BCSSS)
felix.tretter@bcsss.org; Mobil: 0043 660 6270 666

Bertalanffy Center for the Study of Systems Science (BCSSS)
1040 Vienna, Paulanergasse 13 / 2nd floor, door 5

- Workshop I (of a series of workshops) -
10 October afternoon + 11 October morning 2018

Nothing is more practical
than a good theory (K. Lewin).

More data does not mean
more knowledge!

FOREWORD

The final aim of this paper is to provide a later and more widely formulated statement or special publication (e.g. "Memorandum") on the *epistemic relevance of Big Data* (BD). It will therefore initially serve to prepare a further workshop, the focus of which will be on BD applications in the *health sector*. The paper is based on the contributions and discussions of the first workshop and is a kind of perspective paper. The workshops follow the aims of Ludwig von Bertalanffy who criticized over-applications of systems and cybernetic thinking in human affairs (Bertalanffy 1967[1]).

Basically, the paper assumes (1) a socio-political and (2) a "knowledge topological" premise:

1. *Knowledge should be a common property*, because - according to the position of enlightment ("Aufklärung") - science serves the information of the population and the social institutions (Bacon 1620[2], Böhme 1993[3]). Consequently, knowledge asymmetries, and - above all - intransparencies arising from the privatization and information aggregation of BD must be avoided. This position corresponds to our constitutive socio-political value "democracy", which is the political institution of the modern constitutional state and which is the model for most of the world's states. It has a special significance in the context of globalized digitalization of society as a whole (Hofstetter 2016 [4]).
It must be clearly seen, however, that reality is moving in the opposite direction towards the *privatization of "knowledge"* by data corporations, with the *legal expropriation* and creation of a data-based shadow figure as a *representation* (model) of each individual person. In addition, the institutionalization of an elusive *data trade* with an even more elusive *data economy is* emerging in a dark nearly invisible field

---

[1] Bertalanffy, L.v. 1967. Robots, men and minds. Braziller, New York
[2] Bacon, Francis. 1620/1994. Novum Organum. Chicago, Il: Open Court.
[3] Böhme, G. 1993. At the end of the Bacon age: studies on the development of science. Suhrkamp, Frankfurt
[4] Hofstetter, Y. 2016: The End of Democracy - How Artificial Intelligence Takes Over Politics and Disempowers Us, Bertelsmann

of our society. The possible role of (public) *universities* as an institution of *quality assurance of knowledge* - and in that respect also for BD - is understandably, but unfortunately, *too weak* in this development. It is therefore not in the scope of a civil society. However, this complex and highly dynamic social dimension of BD would have to be discussed in a different setting.

The methodologically appropriate discipline for reflecting on the cognitive value of BD is *knowledge philosophy*, based on the traditional *epistemology* (Baumann 2015 [5], Ernst 2016 [6], Gabriel 2012 [7], Andi 2010 [8], Sosa 2017 [9]), *philosophy of science* (Schurz 2014[10]) and the method-reflexive *methodologies of the empirical sciences* and the *formal sciences* (e.g. philosophy of mathematics, logic of statistics, measurement theory) or *system philosophy* (Bunge 1998 [11]). In addition to this inter- and supradisciplinary spectrum of positions, there are insights from *reflective practitioners in* the ICT and BD industries, which constitute the "transdisciplinary" dimension of this project (Mittelstrass 2003 [12], Scholz 2011 [13]).

WORKSHOP

The workshop, which is the subject of this text, pursued the following questions on BD, which were nicely accentuated by Roland Scholz. What is / are:

(1) the "added value“? - This was treated very centrally.
(2) epistemic boundaries? - This was also a priority.
(3) naive overestimations? - Here, too, examples were discussed.
(4) unintended side effects on the science system or knowledge system of society? The latter point has already been mentioned above.

The following paper contains the program text known to the WS participants and the results of the WS as a summary and accentuation of the contributions of Dirk Helbing, Roland Scholz, Yvonne Hofstetter, Christian Brandlhuber, Joachim Mau and Felix Tretter. As a "note" there are some more detailed information and as an "appendix" there is the program, some theses and the affiliations of the speakers.

A. RESULTS OF WS 10 / 11 Oct. 2018

The central question of the WS was to critically assess the *epistemic potential of BD. Above all*, the "successes" of BD communicated in the mass media (and elsewhere) are on the one hand very positive and, because of that unbalanced information, appear mostly as *business-motivated exaggerations*, and also because of the excessive data faith they are dangerously distorting as far as “discovering reality” is

---

[5] Baumann, P.(2015) Epistemology. Metzler, Stuttgart
[6] Ernst, S (2016). Introduction to epistemology. Knowledge. Buchges. Darmstadt
[7] Gabriel, M., (2012) The knowledge of the world. An introduction to general epistemology.
[8] Audi, R. 2010. Epistemology: A Contemporary Introduction to the Theory of Knowledge. Routledge, London
[9] Sosa, E.(2017). Epistemology. Princeton Univ. Press, Princeton
[10] Apron, G. (2017). Philosophy of science. WBG, Wiesbaden
[11] Bunge M (1998). Philosophy of Science. 2 Bde. Transaction Publishers, Piscataway, NJ
[12] Mittelstraß, J., (2003): Transdisciplinarity - scientific future and institutional reality. Constance (Constance University Speeches 214)
[13] Scholz, R. (2011). Environmental Literacy in Science and Society. From Knowledge to Decisons. Cambridge Univ. Press, Cambridge

concerned. In order to make this fundamental skepticism more precise, a philosophically oriented multidisciplinary investigation as an epistemology of BD is required. First steps in this direction were discussed at the WS.

1. Epistemology of Plato's cave parable - data shadow and the "reality

Plato has already clarified the problem of the justification of the truth of knowledge as a "justified true faith". He constructed the cave parable, which was interpreted and changed in various ways in the course of our intellectual history (Plato, after Blumenberg 1989 [14]): People bound in a cave see shadow plays on the wall; one person succeeds in escaping, and he sees that in front of the cave people walk up and down in front of a fire, and he therefore understands that the shadows on the wall are shadows and not "reality". But he can't come back and tell the tied up people about it! Thus it is not possible for humans to recognize the "reality of reality" and therefore the "truths" about the world are not certain but from a constructivist point of view only ”viable”. This insight probably also applies to BD (NOTE 1)

2. Conceptual apparatus and model for epistemic analysis of BD

To study and describe the epistemic-heuristic value of Big Data (BD), a *semantic taxonomy of the terms is* particularly important: If it is assumed that "more data also means more knowledge", then the relevant terminology must be examined critically. This is, above all, due to the necessarily pronounced *interdisciplinarity of such discourses*, which can be centered on philosophical *epistemology*, *theory of science* and *methodology*, but which nevertheless contains many new aspects that go beyond classical epistemology.
It is particularly important to understand 'knowledge' first of all as circular-cyclical *knowledge production, in the* form of a process circuit whose components can be described as follows (Bunge 1998 [15]):

- the object, the *epistemic object*
- the *observation*, i.e. "data".
- the *model of* an object
- the *intervention that* makes the object measurable or changes its conditions

It should be noted that even the *observation* is influenced in multiple ways: not only by the *object* itself, but also by the systems noise, by the *model* that works as a guiding structure for observations, as well as by the *measurement procedure* influence the observation. The latter problem has become known in principle in quantum physics through Heisenberg's “uncertainty relation”. Therefore, in the observation as a whole, falsifications and selections occur that impair the truth value of the observation, and thus its *validity*, but also its *reliability* as a *bias* (Fig. 1). This outline is based on other circular concepts of reality construction as epistemic cycles (NOTE 2).

[14] Blumenberg, H. (1989). *Cave exits*, Suhrkamp, Frankfurt am Main
[15] Bunge M (1998). Philosophy of Science. 2 Bde. Transaction Publishers, Piscataway, NJ

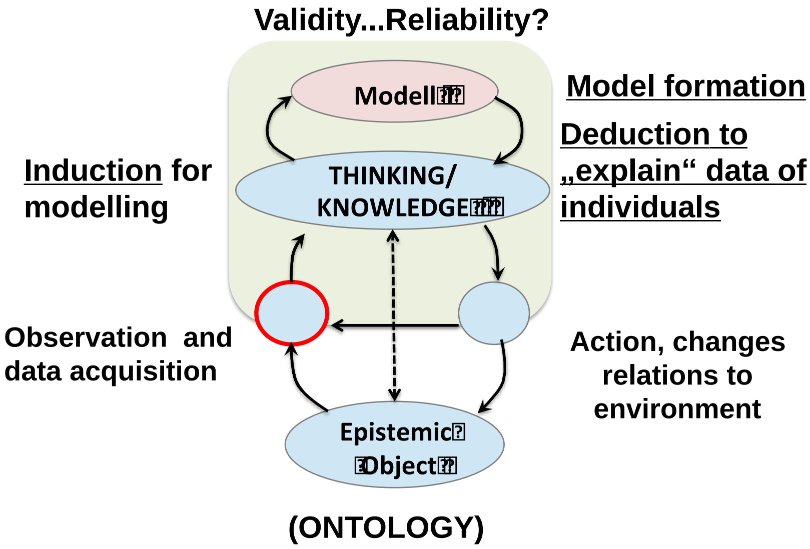

Fig. 1: Knowledge production as an epistemic multi-stage circular process, based on observations and models and measurements as actions, which influence the observation per se, and which also make the object appear in a certain way.

3. The scientific knowledge process and quality criteria

The production of knowledge in modern societies is in the responsibility of science, and this is why, above all, phil*osophy of knowledge* (or more specifically: the *philosophy of science*) is used here as the point of view for analysis. In addition, a current brief characterization of science by philosophy of science according to its goals, the fundamental interplay of empiricism and theory, etc. should be considered (NOTE 3).

If the area of *empirical research to* which BD can be assigned is selected, then the question of the *assessment of the quality of data* and the knowledge derived from it arises. For this discourse, the following terms and their meaning must first be specified: *System state of* the research object, *sign* or *signal* (to noise), *data* as metric, *information*, *recognition*, *knowledge* / belief, *communication*, *intervention*, change of the system state etc. In this sense, it is suggested here that the following distinctions for our inquiry on the use of BD in human and social research could be more precise (Fig. 2):

1. *measurement of* the system takes place via measurement and/or sensor technology by an observer, with the recording of
2. *signs* or *signals* of the system as characteristics which differ from noise (see ECG), which - if they are related to a frame (time) - generate *variables* (frequency of cardiac action), which furthermore can be assigned to numbers (scaling), a procedure that
3. generates (quantitative) *data* on measurements that are subsequently processed, for example via statistics and tests, that
4. serve as *information that is* embedded in
5. *knowledge structures (*e.g. about health/illness) and, in the case of contexts, also within the scope
6. of *theories* which are confirmed or have to be rejected.

The observer who communicates with other observers shapes *collective knowledge*, and also enables
7. *interventions* (e.g. incentives or manipulations in the field or experiments)

and thus, again new measurements have to be applied etc.

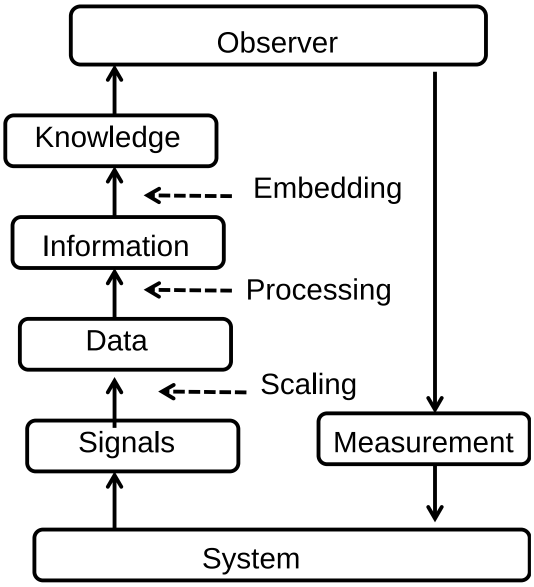

Fig. 2: Epistemic control loop of empirical research

Data thus represent system states by means of methodologically captured signs or signals and must be placed in a conceptual framework that constitutes their meaning in the semantic sense. Typical empirical research is based on scientific theory and also on the assumption that a hypothesis is formulated on a knowledge base, which is examined on a survey instrument (measurement) which uses several samples, whereby various probability-theoretical considerations are applied (distribution assumptions, probability of error, etc.), which allow groupings or delimitations to be made by means of analyses of covariances or variances of the data. These tools of multivariate statistics (MVS) is the basis (or sub-area) of the BD analysis. It is supplemented by *proximity* and *distance measurements*, mostly formulated by graph theory. On general, a special feature of BD is a much more complex data analysis than that provided by MVS.

The central question is: What does Big Data tell us?

BD can establish complex correlations by the extremely fast calculation of the data (*velocity),* their large extent (*volume*) and by the diversity (*Variety*) of the data sets (bio-, behavioral and social data). Thereby, BD is assumed to allow to accurately predict the *specific behavior of humans*, scaled down to single individuals.
The meta-theoretical assumption of BD epistemology, which however is not explicit, can be summarized in the sense of Chris Anderson in such a way that BD needs no theories or hypotheses, because the data would speak for themselves (Andersen 2008 [16]):

> *„Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and*

[16] Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." WIRED Magazine 16/07. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves. ‘‘

These statements, which represent a new pragmatically justified form of "naive empiricism", are, however, not well-founded, because basically known *methodological problems of empirical social research* are not taken into account, one of which is at the center:
If in the sense of the philosophy of science, as described, a main goal of science is to *describe*, *explain*, *understand*, and *predict* its subject as accurately as possible *(Schurz2017*[17], Teten 2013 [18]). In this view, *causality cannot be* clearly derived from *correlations* (or contingencies) based on multivariate data sets (Bortz and Döring 1995 [19], Backhaus et al. 2016 [20]). This can be parodied by the well-known example of the coincidence of the increase of storks and births in such a way that this finding proves that storks bring the babies ("ecological fallacy"). Correlations are only one basis for a causality hypothesis.

5. Methodological and measurement theoretical aspects

BD is the result of ultra-precise measurements in some areas, such as Omics technologies, where, for example, the activity of 22,500 genes can be measured simultaneously. Already this data situation, based on *on-off patterns*, results in a rough estimate of 1.5*10^6773 possible arrangements of different patterns, which appear to be "*transcomputable*": with a hypothetical calculating power of a computer with 10^24 flops for a runtime of approx. 4 billion years (approx. earth age) only approx. 10^41 arithmetic operations would have been realized so far. In this case, the system "genome" is measured dichotomously (on/off), i.e. only nominally scaled. However, if the strength of the gene expression (e.g. its duration) is also taken into account, the hypothetical amount of data becomes even more extensive and thus unmanageable (NOTE 4).
This raises a problem, namely the question of the *optimal granulation of the scales* or the measurement operation, which should represent the interesting features of “reality”: For example, the coastline of England is about 2775 km long when measured with 100 km sections, but about 3425 km long when measured with 50 km sections (Wikipedia 2019 [21]). Using the highest possible resolution of the measurement scales ultimately leads to *disinformation*, for example in the distance estimation when planning a trip to the coast (similar things are generally known from holidays by car!).
This means that the physical and chemical measurement techniques, or the *sensors that* can measure in the nano range, are undoubtedly *highly sensitive* and can acquire a multitude of data (behavior, bio-data, contacts, communications) within the framework of BD, for example via the multi-functionality of mobile phones, without generating any actual added epistemic and / or pragmatic value (D. Helbing).
This problem of the "over-complexity" of the data situation in decisions is further reinforced by the findings of *intuitive information processing* by decision-makers as Kahnemann, Tversky or Gigerenzer have shown (Gigerenzer 2007 [22]).
Important for the interpretation of BD is the *quality of data* and of course also the context of the data source (e.g. registration of noise), and the meanings of the registered communications (D. Helbing).

---

[17] Apron, G. (2017). Philosophy of science. WBG, Wiesbaden
[18] Tetens H. 2013. Philosophy of science. Beck, Munich
[19] Bortz, J., Döring, N. 1995. Research methods and evaluation. Springer, Berlin
[20] Bakehouse, K., Erichson, B., Plinke, W., Weiber, R. 2016. Multivariate analysis methods. Springer, Berlin
[21] Wikipedia Q: https://upload.wikimedia.org/wikipedia/commons/f/f9/Britain-fractal-coastline-50km.png
[22] Gigerenzer, G. 2007. Belly decisions. The intelligence of the unconscious and the power of intuition. Bertelsmann, Munich

From a research-logical point of view, however, the quality characteristics of these high-tech sensing and data processing machines must be clarified: The *signal-to-noise ratio* with much noise can lead to low *specificity at high sensitivity:* A motion sensor in front of doors opens the door automatically, even for people passing too close. Since only a few of the passers-by are also buyers, very many *false positive actions* of this door occur. Thus, when 80% of "correct" diagnoses or forecasts are mentioned, especially in the popular media, it is usually not communicated how often (or as a percentage) *passers-by are wrongly classified as buyer*s. In this respect, the presentation of the *test quality* according to the relationship between sensitivity and specificity (correctly positive hits versus correctly negative hits) is therefore inadequately represented in the public communication of the results of BD. This also concerns problems of *validity*, *reliability* and other topics of *measurement and test theor*y (Bortz and Döring 1995 [23]). In addition, the so-called *bias of* recursive calculations (D. Helbing) must be taken into account: incorrect "learning algorithms" discriminate against women and blacks and *perpetuate these errors* in further iterations.
Dirk Helbing's lecture also showed that *BD can be misleading without theories*: apparently sensational predictions (flu waves) can be explained classically and theoretically by different contact behavior (air traffic). Therefore, in this case, the classical epidemiological compartment model, which distinguishes *susceptible*, *infected* and *resistant* individuals (SIR model), should be *expanded* and not replaced by BD. This already suggests that *theories (and models) are heuristically useful* - contrary to the claims of many BD protagonists.

6. Mathematics by Big Data

With regard to BD, it should be noted that their highly complex data sets are usually analyzed in computerized form (e.g. SPSS and following) according to known principles of *multivariate statistics*, for example via factor, cluster and discriminant analyses, in order to uncover *latent correlations* and *differences* which - intuitively viewed - appear as random noise. Especially in *time series analyses* approaches of nonlinear signal analysis (e.g. chaos theory) have shown progress. Building on this, behavioral prognoses are made, especially in the human area. In addition - for example in the context of *systems biology* and *systems medicine* - *graph theory*, which has been known since Leonhard Euler (!), is used to quantify proximity relationships (in topological significance). Finally, the results are generated by several organizational steps in data processing.
These computational results are obtained in molecular systems biology and now via popular scientific media (e.g. brain research: brain and mind) via the well-known colorful pincushion-like aestheticized visualizations. Especially on this level the question arises also for scientists (but also for laymen) which "knowledge" about "connections" was gained by this method. This is mainly because *no theory* for interpretation (e.g. brain theory) is available (NOTE 5).
In addition, with regard to the biosciences, it should be noted that data do not do fit to the operational principles of living systems, insofar as self-organization, spontaneity, willingness to cooperate, but above all human values such as trust, truthfulness, etc., are regarded as properties that are difficult to measure (D. Helbing). In addition, it would make sense to install alternative models of interaction, namely not war rooms, but peace rooms (Helbing & Seele 2014 [24]).

7. Knowledge

Knowledge is to be regarded as the final stage of knowledge production. It can simply be understood as "justified true faith". It serves for *orientation*, but also for *acting*, it requires beyond that also a *method*

[23] Bortz, J., Döring, N. 1995. loc.cit.
[24] Helbing, D., Seele, P. 2017 Turn war rooms into peace rooms. *Nature* volume 549, page 458 (28 September 2017)

knowledge, how one comes to stable realizations. Knowledge is a product of observation and reflection with the certainty that these cognitions apply to the intended subject area.
Some characteristics or distinctions thereto:
Even *everyday knowledge* is based on *concepts* that guide actions. The basic characteristic of knowledge (including science) is the *(subjective) certainty of knowledge, which*, however, is not present even in banal forecasts. Science as an institution of knowledge production thus ultimately also does not lack "faith". However, it is based on *comparison*, especially according to the cognitive logic of randomized, blinded pair comparison with the search for significant (= unlikely random) differences to generate "cognition". This is, for example, one of the "king's paths" to "safe (evidence-based) knowledge" in medicine.

It should also be noted that knowledge research has also shown that there are different forms of knowledge. Above all, *implicit knowledge* is relevant for action, as decision-psychological studies since Kahnemann, Tversky and Gigerenzer have shown (Gigerenzer 2007[25]). This form of knowledge characterizes the intuition of the experienced practitioner, for example of competent physicians (Fig. 3).

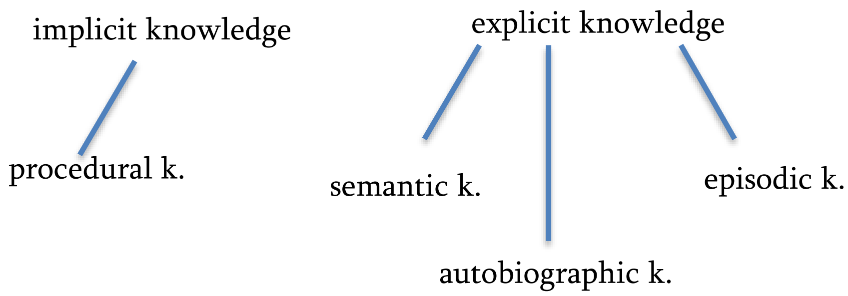

Fig. 3: Forms of knowledge from a psychological point of view, according to degrees of awareness and according to areas of experience and action with correspondence to the structure of memory as a store of knowledge.

> *Martin DUGAS (pers. communic.) - Understanding data!*
> *"From my point of view, we as human beings - especially in times when artificial intelligence is gaining in importance - must continue to claim to be able to understand the algorithms of computers. "Black box algorithms are highly problematic. A purely empirical test of such a system is often impracticable because there are far too many different data inputs with which the black box system would have to be tested. In the public discussion it is often suggested that more data means better data and better conclusions from the data. But that is not the case, rather the opposite. A large data set with one bias in the data collection provides much worse and incorrect results compared to a small data set without bias. Data quality plays a central role - especially in times of big data, because checking data quality in large data sets is correspondingly time-consuming.*

8. Examples of meaningless BD results

[25] Gigerenzer, G. 2007. loc.cit.

Individual examples of highly questionable BD results, as demonstrated in the workshop by Dirk Helbing in particular, show on closer inspection that the analytic algorithms very often produce meaningless correlations that seem to allow "new insights" but do not resist critical empirical analysis by experts. The "ecological fallacy" mentioned above is just one example. However, such method-critical argumentation figures of *inference statistics* are dismissed as inaccurate in the BD discourse, at least in public communication, which often has the character of *marketing by data firms. The* defense of BD in this context is that “it is only in its infancies”, “that we need even faster computers”, and above all even “more data”, including other types of data (e.g. bio-data, genome data).
Thus, classical characteristics of science as a specific *culture of describing, explaining and justifying* are rejected for no reason. In addition, BD results ignore proven *theories* and thus violate the well-confirmed rule that data only make sense in the context of more comprehensive systems of notions and principles (e.g. theories).

Important application areas of BD are questions of the military, the financial economy and the real economy. AI instruments are used to detect patterns in seemingly irregular data matrices. This should be possible via *learning algorithms.* In order to achieve this goal, a large amount of data that does not seem to have anything to do with the core topic is collected and correlations are calculated. The method problem that arises is multi-sensor data fusion, a problem that is still being discussed as knowledge integration.

> *Dirk HELBING - "We have the oil, but the engine is still missing..."*
> "There must be sufficient scientific and ethical competence on board. And an interdisciplinary and multi-perspective approach. The difficulties associated with large amounts of data have been underestimated. These include false correlations, patterns that are found in the data but have no meaning."
> (Q: https://www.bilanz.ch/unternehmen/das-ol-braucht-noch-seinen-motor)

9. Processing of heterogeneous data sets - methodology of interdisciplinarity and transdisciplinarity

The heterogeneous mix of variables from different domains, which the BD methodology intends to process primarily in the health sector, but also in areas such as "smart cities", produces a multitude of factual and technical trivialities and even many risks *of error*. For example, it is not easily possible to bring (quantitative) *scientific data into* a meaningful relationship with (qualitative) *social science data* without so-called *bridge concepts*. For example, the more precise data of seismology must be scaled up to a coarser scale level (see NOTE 4).

There are thus problems of *knowledge integration,* which are caused by the *interdisciplinarity of the questions. In* addition, as has been learned over the decades in research, design and management of socio-ecological systems, "transdisciplinarity" is necessary to improve knowledge performance, which consists in integrating the knowledge of stakeholders or those affected by problems into the scientific discourse. However, this means that qualitative (or semi-quantitative) data will definitely be used (NOTE 6).

In addition, a few definitions from *sustainability research* or *social ecology* should be mentioned here, as this topic is currently most intensively discussed in this context, above all because it regards a matter of links between the social sciences and the natural sciences (Scholz 2011 [26]):

[26] Scholz, R. 2011 loc. cit.

INTERDISCIPLINARITY
..."a mode of research that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or field of research practice." (COSEPUP 2005, S. 26)

TRANSDISCIPLINARITY
The International Council for Science (ISCU, 2010, p. 6) states that in order to address sustainability issues comprehensively, a science is needed that integrates expertise from the social, natural, health, engineering and human sciences.

Scholz 2011 [27]:

Knowledge integration and transdisciplinary processes are key strategies for understanding complex human-environment systems.

Transdisciplinarity, according to R. Scholz can also be divided into horizontal (I) and vertical (II) transdisciplinarity (Scholz 2011 [28]).

Continuing the importance of transdisciplinarity as a form of knowledge production, one can also regard the "wisdom of the many" as more advanced than it might be possible for a central *super-intelligence.* Due to the variety of combinations of data and the associated length of the calculation time, this will hardly be technically feasible. Last but not least, *variation* is important for evolution, which runs against standardization and averaging.
A particularly concrete problem can be represented by the *algorithmization of law*: "Equal law for all" suggests that there must be general algorithms, but "proportionality" or "mitigating circumstances" also refer to the comprehensible *individual variation of* judgments. In addition, the legal framework has to be reinterpreted once in a while, because of new reference judgments. Nevertheless, this topic still remains open in the discourse of the BD or AI.

10 The Perspective - Enhancing the Importance of Theories: Example health and disease

An example of the usefulness of theories even in areas where elaborate and empirically-based *theories or explanatory models are* lacking, such as in medicine, psychology or sociology, can be shown, at least with the aid of *general system theory*, that, among other things, *dynamic equilibria* are important for living systems. Here are some examples based on the *asymmetric, fluctuating* and *adaptive equilibrium of activation and inhibition* in the organism:

- the autonomic nervous system, consisting of two counterparts, the *activating sympathetic nervous system* and the *calming parasympathetic nervous system* (also called vagus), regulates the activity of the various organs, also in circadian rhythm: the sympathetic nervous system increases the *heart rate*, while the nervus vagus decreases it, etc. The obviously *antagonistic innervation of* the organs is neither anatomically nor physiologically symmetrical, but to a certain extent *asymmetrical*, otherwise there would be a standstill. One can imagine this structure of influence as a balance oscillating in a slight obliquity, whereby this obliquity is reversed in the circadian course - in the morning, a dominance of the sympathetic nervous system, at night dominance of the parasympathetic nervous system occurs.

[27] Scholz, R. 2011. A.i.o.
[28] Scholz, R. 2011. A.i.o.

We have used this concept of multiple neurochemical antagonisms for six neurotransmitters of the brain within a "neurochemical mobile" to model the neurobiology of mental disorders (Qui et al .2014 [29]).

- the regulation of the *stress system*, which influences all organ functions from the hypothalamus, the pituitary gland and the adrenal cortex via cortisol (Sriram et al. 2012 [30]) and which furthermore converges retroactively via feedback circles on several levels of the brain.

- The regulation of *blood flow properties is* achieved by a demand-driven adaptation of the many molecular factors that physiologically create a balance between *blood clotting in* vascular injuries and *fibrinolysis* for the natural dissolution of small fibrin deposits in vessels. Deficiencies of one or the other component with its many molecular (and cellular) factors of haemostasis or dissolution of blood clots lead to difficult to stop bleeding or to an increased tendency to thrombosis. These deficiencies can be substituted by medication (i.e. antagonized by antidotes), such as coumarins as vitamin K antagonists for prophylaxis in cases of high risk of thrombosis, e.g. after a heart attack.

Since there are generally *several cascades of mutual inhibitory factors* or cellular elements in different areas of the organism, especially at the molecular level, and especially in the brain, which in the final consequence at the *effector* cause an activation, the measurement of the concentration (or activity) of these factors without a model conception of the modes of action is not very target-oriented, especially if one does not identify the individual members: If A inhibits B and B also inhibits the element C, then at a high concentration of A the concentration of C is also high and therefore A is wrongly classified as an activator instead of a *disinhibitor of* C. This also applies to other biochemical mechanisms in the cell, as a *sequential cascade of inhibitions* is very often identified.

If now also *feedback loops* in the organism are considered, especially when it comes to *delayed feedback*, then measurements of individual factors would show almost chaotic distributions. It is therefore easier to understand the functions of the organism by identifying the *causal relationships that* often function as *control loops in which* a certain measured value fluctuates around an average value (e.g. heart rate, blood pressure, thyroid hormones, blood sugar, cortisol, etc.). Since the organism consists of a *multitude of coupled control loops,* a consistent cybernetic modeling of the organism as a *system of systems is* useful. Mr. Joachim Maus pointed this out in our workshop.

If, however, one tries to simulate *more complex networks* consisting of activating and inhibiting elements on the computer, very complex and unpredictable behavior patterns result. Systems biology is therefore attempting to use *mathematical graph theory to* identify simple *canonical modules that*, as *building blocks of larger networks*, can be used to better understand the behavior of the entire body of work. However, it is remarkable that only a few researchers like Uri Alon in systems biology are pursuing this path (Alon 2006 [31]).

> *Jochen MAU - Relationship Data / Practice / Theory*
> - "The statistical significance of beautiful appearance." This is connected with an increasingly articulating rejection of "meta-analyses = quantitative research syntheses" as quite nonsense -> goes completely wrong for medical practice (individual therapy).
> - As corollary, there is a very critical attitude to guidelines: I had advised my students to develop their medical intuition instead of relying on guidelines blindly (!) [which then results compulsively - "there is a guideline after all"].

[29] Qi Z, Tretter F, Voit EO. 2014. A heuristic model of alcohol dependence.
PLoS One. 2014 Mar 21;9(3):e92221. doi: 10.1371/journal.pone.0092221. eCollection 2014.

[30] Sriram, K., Rodriguez-Fernandez, M., & Doyle, F. J. (2012). Modeling cortisol dynamics in the neuro-endocrine axis distinguishes normal, depression, and post-traumatic stress disorder (PTSD) in humans. *PLoS computational biology*, *8*(2), e1002379.

[31] Alon, U. 2006 An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall. Boca Raton

- "Use assistance systems generally only when you really need them; if you use them constantly for convenience, make yourself incompetent / stupid!"
- Biocybernetics is essential for medicine (see 2018 December Conference in Düsseldorf)

11. CONCLUSION

It is a matter of a reasonable assessment of the epistemic relevance of the data collection, which, without theory development, quickly becomes a cryptic project that is not only unrealistic, but that can actually lead to a damage through associated wrongly founded actions.
Beyond that it concerns - as well-known - the responsible use of data, considering the advantages and disadvantages, a procedure that is demanded for example in the project DiDat.

In addition some differentiated theses of Roland Scholz are mentioned (project "Unintended effects of Big Data"; "DiDaT"):

1. Big Data analytics is/will become a major tool of power in society
2. There are 'haves' and 'haves not' Digital Data divide society with respect to
* big data access
* Big data analytics (cloud computing might change)
3. Big Data analytics is a matter of private and governmental knowledge institutions and not of universities
4. Access to Big Data is a matter of ownership and payment or "autocratic governmental access"
5. The "knowledge behind" (hacker knowledge, getting access), i.e., how to get access to Big Raw Data is more important than the knowledge how to analyze.
6. Big Data analytics is a primary weapon of business and political leadership (Political and economic surveillance society)
7. Imbalance/injustice for those who become subject of Big Data analytics
8. A comprehensive, sustainable 'rule system' of how Big Date should be used by whom in what way is missing
9. Big Data and computational science (also when looking at "third mission variants") will fundamentally change the role of universities as "sustainability science" has to become a lead science and Sustainable Digital Environments shall become a major subject of sustainability science

12. RESUME

Felix TRETTER - simple theoretical models to reduce complexity are wanted!
- Big data in the current organizational form finalizes public science in the sense of Sir Francis Bacon and thus democracy (see Anticipations of Gernot Böhme "Finalization" of Science 1974).
- False truths" can therefore be collectively propagated in an unverifiable way (marketing rhetoric instead of exact - also understandable - scientific language).
"New" is the totalitarianism of the scientific organization: sample size has always been the limiting factor in the "truth-finding" of empirical research.
- The quality of measurement and test theory is worse than the public representation by the protagonists.
- The explicit epistemological and scientific-theoretical location of big data is missing.

NOTE 1

This form of (solipsistic) idealism can be found in a similar form later in constructivism (especially by Glasersfeld 2000 [32]). *Constructivism*, in fact, classified the question of certainty about the existence and form of "reality" as insoluble and therefore as subordinate and conventionalist. Nevertheless, a difference between landscape and map is undeniable, insofar as a map allows its constructor more or less "viable" actions. This view is similar to the traditional *critical rationalism of* K. Popper with, among other things, the *falsification principle* of the resilient discovery of truth (Popper 2002 [33]). Collective knowledge is thus ultimately a social construction of reality.

NOTE 2

In a further subdivision of the *production of expert knowledge it* can be shown that knowledge generation about *practitioners* makes the observations in their field of action, *empirical research that* generates data and sorts findings via *theory*, can be demonstrated. This has already been shown in detail on medicine (Tretter 1996[34]; Fig. 7): Qualitative observations in the clinical context (1), as in investigations of new diseases (e.g. most recently AIDS), lead to new findings for the understanding of which initially existing theories are used (2a), but at the same time empirical studies are stimulated to quantify the phenomenon (2b). These data from empiricists (infectiologists, epidemiologists) are evaluated using formal theories, whereby various discrepancies are bridged by theorems. The empiricists can then make practice-relevant scales available to the practitioners, and so on.

In medicine, a special form of securing knowledge about therapeutic intervention effects has established itself, namely *evidence-based medicine (EBM)*, which refers to *the research logic of randomized controlled trials*: these are randomized, blinded pair comparisons with the search for significant (= improbably random) *differences* as logical-empirical proof of effectiveness of verum interventions versus placebo interventions.

[32] Glasersfeld , E.v. 2000. Radical Constructivism : A Way of Learning. Taylor & Francis, London

[33] Popper, K (2002) [*The Logic of Scientific Discovery*. Abingdon-on-Thames: Routledge

[34] Tretter, F. 1996 Wissenschaftstheorie – Umweltmedizin: Beschreibungen sind derzeit wichtiger als Erklärungen. Dtsch Arztebl 1996; 93(34-35): A-2136 / B-1812 / C-1704

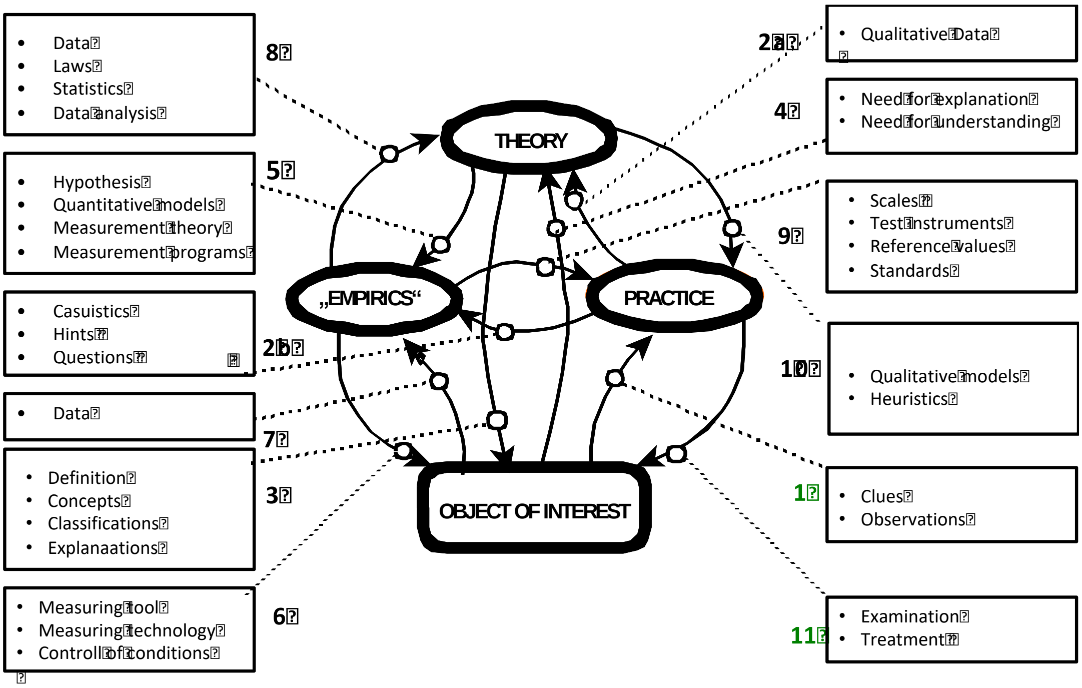

Fig. 2: Knowledge process in medicine and the theory-practice relationship (Tretter 1996).

NOTE 3

<u>Scientific objectives: (cf. Tetens, 2013 , p. 17).</u>

a) Truth:
Exclusive access to reality.

(b) Justifiability:
The statements should be based on observations and argumentation according to the logical rules.

c) Explain and understand:
Empirical laws are established from a large number of individual observations, which allow new observations to be explained or even make forecasts possible.

(d) Intersubjectivity:
By specifying the conditions of the observations or procedures by which considerations are made to ensure traceability.

e) *Self-reflection:*
The opportunity to occasionally think about scientific activity. This concerns epistemology or the theory and philosophy of science. Such reflections on the consequences of human cognitive programs can be found in physics: For example, Heisenberg's uncertainty relation shows that location and impulse are not determined exactly at any time, but only to a limited extent. It was finally Albert Einstein who said: ".... Science without epistemology is primitive and confused ...".

NOTE 4

Measurement is therefore the assignment of characteristics of the object under investigation (*empirical relativs*) to elements of the level of numbers (*numerical relativs*). The distances or differences of the characteristics can be brought into an order which - if a series can be formed, like those according to size (*larger* or *smaller*) - forms a ranking (*ordinal scale).* If the differences are the same, it is an *interval scale*. The interval scales can be subdivided into millimeters or kilometers etc., i.e. they can be based on different units of measurement.

NOTE 5

Fundamentally, it must be made clear that the *function of mathematics in* data analysis is *interpolation*, or *extrapolation* based on the data sets, and thus the *optimal approximation to empirics takes* place via equations ("line of best fit").
With regard to the associated claim to forecasting competence, it should also be noted that the weather forecasts fed with a large amount of physical data in the sense of BD have become weaker again in their forecasting power due to the poorly modeled local weather dynamics, which have recently become more relevant (Latif 2009 [35]).

NOTE 6

With regard to knowledge integration in the sciences, a distinction must first be made between "associative" and "integrative knowledge integration" (and intermediate or special forms taking complementarity, overlap, etc. into account differentially). Especially interesting is the connection between the *measuring natural sciences* and the *understanding social sciences,* insofar as it does not seem to have been conclusively made today. Although physics can "integrate" chemistry in the form of the physics of the outer electron orbits and thus achieve a theoretical reduction according to E. Nagel, the complete reduction of biological phenomena has by no means succeeded. Especially when you think of solid *explanations of* life phenomena as well as consciousness phenomena.
Analogies are often used in this context *as heuristics*, a strategy which is also used by the "supradisciplinarily" applicable structural sciences, e.g. *mathematics*, but also *system theory*: Terms like *structure, dynamics* etc. are abstract and applicable across all disciplines. On the other hand - and this is important in methodologically heterogeneous disciplines such as social ecology with different system levels (geo-, bio-, physico-chemical etc. sphere) - *specific disciplines* have specific and also metricized *terms*. These are room size in m3, weight in kg, temperature, water level, dollars/day/ ^{m2}, percentage of belonging to an educational class, persons /m2 etc. The bridge between *economics* and *sociology is* also difficult to build: how can price-quantity relations, which are determined within the framework of categories of costs and benefits, supply and demand, etc., be "integrated" into sociological categories, for example, or simply "associated"? How can "Good life for all" be quantified, for example by monetarized happiness measurement? So in this case - for example by BD - everything can ultimately be "monetarized" using complex algorithms?
How can the measured variables be combined? Can *quantitative methods of* measuring the environment, for example in the geosciences, but also in economics and *qualitative methods* (interviews) in the social sciences, be combined in a formal approach? m^2 usable area for agricultural economics with meat yields and revenues of x € / kg according to varying market prices (supply / demand) and consumer attitudes towards animal ethics, ecology, health effects etc. of meat - all this and more would have to be reduced to a common denominator!
LITERATURE (important work of the speakers)

[35] Latif, M. 2009 Climate change and climate dynamics. UTB, Stuttgart

- Dugas, M. (2017). Medical Informatics. A compendium for study and practice. Springer, Berlin
- Helbing, D. (2015). The Automation of Society is Next: How to Survive the Digital Revolution. CreateSpace Independent Publishing Platform
- Helbing, D. (2018). Towards Digital Enlightenment: Essays on the Dark and Light Sides of the Digital Revolution. Springer, Berlin
- Hofstetter, Y. (2014). They know everything. - How intelligent machines invade our lives and why we have to fight for our freedom, Bertelsmann
- Hofstetter, Y. (2015). The power of algorithms : self-determination despite(t) despite(t) artificial intelligence, Akademie für Politische Bildung, Bayerischer Landtag, 11 November 2014, Munich / Tutzing
- Hofstetter, Y. (2016). The End of Democracy - How Artificial Intelligence Takes Over Politics and Disempowers Us, Bertelsmann
- Mainzer, K. (2014). The calculation of the world. From the world formula to Big Data. Beck, Munich
- Mainzer, K. (2016). Artificial intelligence, - When do the machines take over? Springer, Berlin
- Tretter, F. (2005). Systems theory in the clinical context. Pabst, Lengereich
- Tretter, F. et al. (Ed)(2010)(2010). Systems Biology in Psychiatric Researh. Wiley, Weinheim
- Tretter, F. (2017) Homo Informaticus: Image of Man in Information Society: Social and Technological Problems. In: Hofkirchner, W., Burgin, M. (eds) The Future Information Society. World Scientific. Singapore, pp.475-505 10.1142/9789813108974_0027.

GENERAL LITERATURE LIST

Bechtel W (2006) Discovering Cell Mechanisms: The Creation of Modern Cell Biology. Cambridge, UK: Cambridge University Press.

Callebaut, W. 2012. "Scientific perspectivism: A philosopher of science's response to the challenge of big data biology." Studies in History and Philosophy of Biological and Biomedical Science 43(1):69-80.

Cartwright, N. 1983. How the Laws of Physics Lie. Oxford: Oxford University Press.

Craver CF and Darden L (2013) In Search of Biological Mechanisms: Discoveries across the Life Sciences. Chicago, IL: University of Chicago Press.

Dupré, J (2012) Processes of Life. Oxford, UK: Oxford University Press.

Floridi, L. (2012): Big Data and Their Epistemological Challenge. Philosophy & Technology 25(4).

Frické, M. (2015): Big Data and its epistemology. Journal of the Association for Information Science and Technology 66(4).

Greeen, S. (Ed) . Philosophy of Systems Biology. Springer, Berlin

Hacking, I. 1983. Representing and Intervening. Cambridge: Cambridge University Press.

Hey, T., Tansley, S., and Tolle, K. (Eds.) (2009): The fourth paradigm: Data-intensive scientific discovery. Redmond,WA: Microsoft Research.

Kelling, S., Hochachka, W., Fink, D., Riedewald, M., Caruana, R., Ballard, G., and Hooker, G. (2009): Data-intensive Science: A New Paradigm for Biodiversity Studies. BioScience 59(7).

Kitchin, R. (2014): Big Data, new epistemologies and paradigm shifts. Big Data & Society 1(1).

Kitchin, R. and McArdle, G. (2016): What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society 3(1).

Leonelli, S. 2012. "Classificatory Theory in Data-Intensive Science: The Case of Open Biomedical Ontologies." International Studies in the Philosophy of Science 26(1):47- 65.

Leonelli, S. (2014): What difference does quantity make? On the epistemology of Big Data in biology. Big Data & Society 1(1).

Leonelli S, Smirnoff N, Moore J, Cook C and Bastow R (2013) Making open data work in plant science. Journal for Experimental Botany 64(14): 4109–4117.

Mayer-Schönberger, V., Cukier, K(2013) BigData: A Revolution That Will Transform How We Live, Work and Think. London: John Murray Publisher.

Mitchell, S. 2003: *Biological Complexity and Integrative Pluralism*, Cambridge: Cambridge University Press

Mitchell, S. 2009: Unsimple Truths: Science,Complexity and Policy, Chicago: University of Chicago Press,

Petersson, G.J. , Breul, J. D. (Ed) 2017. Cyber Society, Big Data, and Evaluation: Comparative Policy Evaluation. Transaction , New Brunswick

Pietsch,W., Wernecke,J. Ott, M. 2017. Predictability of the world? Philosophy and science in the age of Big Data. Springer, Berlin

Psillos, p. 2002. Causation and Explanation. Durham: Acumen.

Rheinberger H-J (2011) Infra-experimentality: From traces to data, from data to patterning facts. History of Science 49(3): 337–348.

Russo, F. 2009. Causality and Causal Modelling in the Social Sciences. Measuring Variations. Springer.

Woodward, J. 2003. Making Things Happen: A Theory of Causal Explanation. Oxford: Oxford University Press.

- -

<u>PROGRAMM (REFERATE mit ARBEITSTITELN)</u>

MITTWOCH, 10. Oktober 2016

13 .00 Begrüßung

13.15 EINFÜHRUNG
Felix TRETTER (Wien)

13.45 BIG DATA – SOZIALWISSENSCHAFTLICHE PERSPEKTIVE
Dirk HELBING (Zürich)

14.15 Diskussion

14.45 Pause / informelle Gespräche

15. 00 BIG DATA – BILATERALE TRANSDISZIPLINARITÄT MÖGLICH ?
How does Big Data Analytics change science and the power relations among stakeholders – transdisciplinary perspectives
Roland SCHOLZ (Zürich)

15.30 Diskussion

16.00 SYSTEMS MEDICINE – MOLECULAR BIG DATA VON GESUNDHEIT UND KRANKHEIT
Felix TRETTER (Wien)

16.30 Diskussion

17.00 GESAMTDISKUSSION

18.00 Ende

DONNERSTAG, 11.Oktober 2016

9:00 ÜBERLEITUNG VOM VORTATG
Felix TRETTER (Wien)

9.30 BIG DATA - IN DER PRAXIS
Yvonne HOFSTETTER (München)

10.00 Diskussion

10.30 Pause

10.45 BIG DATA IN DER GESUNDHEITSSYSTEMFORSCHUNG
NIKI POPPER (Wien) ?

11.15 Diskussion

11.30 TRANSKULTURELLE THEORIEPERSPEKTIVEN DER MEDIZIN – BIOKYBERNETIK
Jochen MAU (Düsseldorf)

12.00 Diskussion

12.30 Zusammenfassung

13.00 ENDE DES WORKSHOPS

EINIGE THESEN DER REFERENTEN / TEILNEHMER (alphabetisch, *Hervorhebung von FT):*)

Martin DUGAS (am 10./11.Okt. verhindert) - Daten verstehen!:
Aus meiner Sicht müssen wir als Menschen - gerade in Zeiten, wo künstliche Intelligenz an Bedeutung gewinnt - weiterhin den Anspruch erheben, die Algorithmen der Computer verstehen zu können. "Black Box"-Algorithmen sind hochproblematisch.
Eine rein empirische Prüfung derartiger System ist häufig unpraktikabel, weil es viel zu viel verschiedene Daten-Inputs gibt, mit denen das Black-Box-System getestet werden müsste.
In der öffentlichen Diskussion wird oft suggeriert, dass mehr Daten bessere Daten und bessere Schlüsse aus den Daten bedeuten. Das ist aber nicht der Fall, eher im Gegenteil.
Ein großer Datensatz mit einem Bias in der Datenerhebung liefert viel schlechtere und falsche Ergebnisse im Vergleich zu einem kleinen Datensatz ohne Bias.

Die Datenqualität spielt eine zentrale Rolle - gerade in Zeiten von Big Data, weil die Überprüfung der Datenqualität in großen Datensätzen entsprechend aufwändig ist.

Dirk HELBING – „Wr haben das Öl, aber die Maschine fehlt noch...“

„Es braucht genügend wissenschaftliche und ethische Kompetenz an Bord. Und einen interdisziplinären und multiperspektivischen Ansatz. Man hat die Schwierigkeiten unterschätzt, die mit grossen Datenmengen einhergehen. Dazu zählen Scheinkorrelationen, also Muster, die man in den Daten findet, die aber keine Bedeutung haben.“
(Q: https://www.bilanz.ch/unternehmen/das-ol-braucht-noch-seinen-motor)

Jochen MAU – Verhältnis Daten / Praxis / Theorie

- "Der statistischen Signifikanz schöner Schein". Verbunden damit ist meine sich zunehmend artikulierende Ablehnung von "Meta-Analysen = quantitative Forschungssynthesen" als ziemlichen Un-Sinn -> geht für die ärztliche Praxis (individual-therapeutisch) völlig daneben.
- Als Korollar ergibt sich eine sehr kritische Einstellung zu Leitlinien - meinen Studenten hatte ich angeraten, ihre ärztliche Intuition zu entwickeln, anstatt sich auf Leitlinien etwa gar blind (!) zu verlassen [das resultiert dann zwanghaft - "es gibt doch eine Leitlinie"]
- "Assistenzsysteme generell nur benutzen, wenn man sie wirklich braucht; wer sie ständig aus Bequemlichkeit benutzt, macht sich inkompetent / blöd!"
- Biokybernetik ist für die Medizin essentiell (s. Dezember-Tagung in Düsseldorf)

Roland SCHOLZ – Transdisziplinarität ist für das Verstehen komplexer Systeme erforderlich!

- Wissensintegration und transdisziplinäre Prozesse sind Schlüsselstrategien für das Verständnis von komplexen Mensch-Umwelt Systemen.

Felix TRETTER – simple theoretische Modelle zur Reduktion der Komplexität gesucht!

- Big data in der jetzigen Organisationsform finalisiert die öffentliche Wissenschaft i. S. von Sir Francis Bacon und damit die Demokratie (vgl. Antizipationen von Gernot Böhme „Finalisierung“ der Wissenschaft 1974)
- „falsche Wahrheiten“ können daher unüberprüfbar kollektiv propagiert werden (Marketing-Rethorik statt exakte (auch verstehbare) Wissenschaftssprache
-„neu“ ist der Totalitarismus der Wissenschaftsorganisation: Stichprobengrösse war immer der limitierende Faktor bei der „Wahrheitsfindung“ der empirischen Forschung
- die mess- und testtheoretische Qualität ist schlechter als die öffentliche Darstellung der Akteure
- es fehlt die epistemologische und wissenschaftstheoretische Einordnung