Efstathiou*, Sophia; Nydal, Rune; Lægreid, Astrid; Kuiper, Martin
Scientific knowledge in the age of computation: Explicated, computable and manageable?

THEORIA

Scientific knowledge in the age of computation: Explicated, computable and manageable?

(El conocimiento científico en la era de la computación: ¿explicado, computable y manejable?)

Sophia Efstathiou*, with Rune Nydal, Astrid Lægreid, Martin Kuiper

Norwegian University of Science and Technology

ABSTRACT: With increasing publication and data production, scientific knowledge presents not simply an achievement but also a challenge. Scientific publications and data are increasingly treated as resources that need to be digitally 'managed.' This gives rise to scientific Knowledge Management (KM): second-order scientific work aiming to systematically collect, take care of and mobilise first-hand disciplinary knowledge and data in order to provide new first-order scientific knowledge. We follow the work of Leonelli (2014, 2016), Efstathiou (2012, 2016) and Hislop (2013) in our analysis of the use of KM in semantic systems biology. Through an empirical philosophical account of KM-enabled biological research, we argue that KM helps produce new first-order biological knowledge that did not exist before, and which could not have been produced by traditional means. KM work is enabled by conceiving of 'knowledge' as an object for computational science: as *explicated* in the text of biological articles and *computable* via appropriate data and metadata. However, these founded knowledge concepts enabling computational KM risk focusing on only computationally tractable data as knowledge, underestimating practice-based knowing and its significance in ensuring the validity of 'manageable' knowledge as knowledge.

KEYWORDS: Knowledge Management, systems biology, cellular signalling networks, knowledge concepts, objectivist epistemology, practice-based epistemology, founded concepts

RESUMEN: Con el aumento de la publicación y la producción de datos, el conocimiento científico no solo es reconocido como un logro, sino también como un desafío. Las publicaciones y los datos científicos se tratan cada vez más como recursos que deben ser 'gestionados' digitalmente. Esto da lugar a la Gestión del Conocimiento científico (Knowledge Management (KM)): labor científica de segundo orden destinada a recopilar, cuidar y movilizar de forma directa el conocimiento disciplinario de primera mano y los datos para proporcionar nuevos conocimientos científicos de primer orden. Seguimos el trabajo de Leonelli (2014, 2016), Efstathiou (2012, 2016) y Hislop (2013) en nuestro análisis del uso de la KM en la biología de sistemas semánticos. A través de una descripción filosófica empírica de la investigación biológica habilitada para KM, argumentamos que KM ayuda a producir un nuevo conocimiento biológico de primer orden que no existía antes y que no podría haber sido producido por medios tradicionales. El trabajo de KM está facultado para concebir el "conocimiento" como un objeto para la ciencia computacional: como algo explicitado en el texto de artículos biológicos y como computable a través de datos y metadatos apropiados. Sin embargo, los conceptos fundados permiten el riesgo computacional de KM de centrarse solo en los datos que se pueden tratar de manera computacional como conocimiento, subestimando el conocimiento basado en la práctica y su importancia para garantizar la validez del conocimiento "manejable" como conocimiento.

PALABRAS CLAVE: Gestión del conocimiento, biología de sistemas, redes de señalización celular, conceptos de conocimiento, epistemología objetivista, epistemología basada en la práctica, conceptos fundados.

Introduction

Scientific knowledge in the 21st century is not only an achievement but increasingly a challenge. What looks like a great resource—so many publications, so much data—is only a resource if one can manage to manage it—or so scientific Knowledge Management practices propose. The last few decades have witnessed the growth of a meta-level of scientific work: "Knowledge Management" (KM) develops second-order scientific work, geared to collect, take care of and discover first-order scientific knowledge and data, by computational means. How does current second-order KM shape first order scientific knowledge? We answer by considering the case of KM-enabled systems biology.

We expand on the work of Sabina Leonelli on data-centric biology (2014, 2016), Sophia Efstathiou on technical, founded concepts (2012, 2016) and Donald Hislop on organisational Knowledge Management (2013) to argue that, in the case of systems biology, scientific KM is helping to produce new first-order biological knowledge that did not exist before, and which could not have been produced by traditional means. This happens by conceiving of 'knowledge'[1] as an object for computational science: as *explicated* in written text and rendered *computable* via data and appropriate metadata. However, the founded concepts enabling computational KM come at a cost. They risk focusing on only computationally tractable data as knowledge, underestimating practice-based knowing and its significance in ensuring the validity of 'manageable' knowledge. We conclude by reflecting on what a practice-based epistemology in KM would imply, looking to organisational Knowledge Management theory as a guide.

Our thesis derives from joint work among philosophers, biologists and bioinformaticians at the Norwegian University of Science and Technology (NTNU). Our work was funded as an "integrated" interdisciplinary project to investigate Ethical, Legal and Social Aspects of systems biology (cf. similar work in Stegmaier 2009; Rabinow and Bennett 2009; Leonelli 2010. Our own approach is outlined in Nydal *et al.* 2012). From September 2011 to December 2014, the co-authors worked through monthly meetings, the 18-month embedded research of Efstathiou in the lab of Lægreid, co-authorship, text-based reflection and discussion, joint international research trips and conference organisation. Philosophical research used empirical qualitative methods including participant observation, in-depth interviews with fourteen scientists, six of which directly inform this paper, as well as several informal interviews, analyses of scientific texts and of our own co-authored texts (cf. Wagenknecht *et al.* 2015; Van der Burg and Swierstra 2013). While accepting that some critical interests of socio-humanists can become troubled and trouble the frame of a shared research project (Balmer *et al.* 2015), we argue in form and message for practice-based, integrated work as a means to understand scientific knowledge production in the 21st century.

Our paper has three main sections. Section 1 outlines scientific KM and its tools, as second-order scientific work in biology, operating on first-order biological knowledge. Section 2 illustrates the development of new first-order biological knowledge through sec-

---

[1] We use double quotes to mark "words", single for 'concepts' and no quotes for the things they refer to/pick out. Double quotes can also function as "scare-quotes" to mark concepts in need of further analysis.

ond-order KM tools: building a "knowledge assembly" model within the field of systems biology. We reflect on the founded knowledge concepts and epistemologies that drive computational KM in Section 3.

1. Knowledge as a Challenge: Second-order computational Knowledge Management in the life sciences

Derek J. de Solla Price—famous for his idea of 'big' science—reached his conclusion using the rate of scientific publication as a proxy for the growth of science (Price 1961, 1963). The growth of scientific publication is today emerging as a scientific challenge itself. Publication is growing at exponential rates across traditional outlets like journals, and new outlets like open archives, proceedings and home pages, with databases archiving this information struggling to keep up (Larsen and von Ins 2010, 576-600). Digital data has now inherited the sceptre of 'bigness' from science: in 2013, 90% of the world's data had been produced in the last two years (SINTEF 2013). Big data includes data produced by scientific activities, such as high-energy (big) physics, but also and importantly the digital footprints of personal and professional lives lived online. *Data science* is an emerging catch field devising new ways to learn from such digital data (cf. the term's first usage by Cleveland 2001). These new approaches to knowing through publications and data are heavily reliant on computational, quantitative methods and statistical analysis. However, the study of knowledge as a usable resource developed first as a field in the social sciences, as a part of business management and organisation studies.

Knowledge Management (KM) became a focus for organisation studies roughly in the mid 1990s, at the same time as the Internet was becoming popularised and computers cheaper (Hislop 2013). The general focus of organisational KM was how to take care of the knowledge of a corporation: this included developing theoretical understandings of what 'knowledge' is for/in organisations and ways to cultivate, share or otherwise capitalise on this kind of resource. Organisational KM thus spanned epistemological theoretical work, qualitative social science methods such as organisational ethnography, and technically oriented sub-fields, such as utilising Information and Communication Technologies to retain, analyse or share employees' knowledge, in their absence. Though Organisational KM is not a standalone discipline, it is pursued using different disciplinary approaches.

In life science, KM is synonymous to this last type of computational or digital KM. Its methods are more akin to computer science and informatics ones than to social science ones, focusing on the computational management of scientific knowledge. Humans are crucial participants in KM, yet the recruitment of computers is an organising goal. Consider some standard tools developed for second-order KM work on bioscience knowledge (Antezana *et al.* 2009, 393-394).

— *Knowledge Representation* (KR) languages: These are formalisms aimed to represent real-world entities and the relationships between them through abstraction, in the form of logical statements that are computationally comprehensible. KR languages provide "commitments" for how to observe a domain and how to reason over it. Formal KR helps structure communications between different computer systems to avoid ambiguity, for instance when collecting and sharing data.

For such "interoperability", systems need to adopt a shared syntax (a way of parsing entities) and a means of understanding semantics (the assigned meaning) associated with the syntax.

— *Ontologies*: Ontologies can be imagined as taxonomies of what "exists"—really, if one is a realist, or specifically for a particular domain following more pragmatic, pluralist or anti-realist approaches (Chalmers 2009; Lord and Stevens 2010).[2] In biology, "bioontologies" are built to be amenable to computational usage: they are structured through prescribed relations between entities, for example "is_a" or "part_of", using KR languages. They may be understood as vocabularies with a specification of intended meaning, or as "controlled vocabularies" plus relations. Ontologies may be formal, using description logics, or non-formal, when describing meaning in 'natural' language.

— Ontologies are populated with information through *curation* or *biocuration*. Curation involves "extracting knowledge" from text and is done usually "manually", i.e. by people (Antezana *et al.* 2009, 394). Biocurators are biology experts engaged in reading the published biological literature and to translate key findings in the scientific literature to annotations of biological entities using expressions composed of terms provided by controlled vocabularies and ontologies, which can then be handled by KR models. Biocurators currently do most of the difficult and uncertain "interpretation" of text (Efstathiou field notes, European Bioinformatics Institute visit -February 6-8, 2013; Leonelli 2014). Biocurators are also often female, employed temporarily, and undervalued (cf. Gabrielsen 2018). Even though demand for biocuration is huge, biologists are not motivated to pursue this work as it is considered less innovative.[3] KM tools are being developed for biocurators to semi-automate information-mining and information-entry—though the prospects of fully replacing humans here is highly unlikely.

— The *Semantic Web* is envisioned as a next generation web that will help computers "interpret" online content. This interpretation will happen, roughly speaking, through extra layers of information. For instance, while reading a Wikipedia article on "cells" you will know from the discursive context whether these are prison cells or eukaryotic cells. A layer of meta-data can make this distinction clear to a computer, for instance by identifying terms through Internationalised Resource Identifiers (IRIs). This is like teaching someone a language by pointing: prison "cell" would be hooked onto a different IRI than biological "cell"; and so too with terms for relations, processes, conjunctions, etc. Standard information exchange languages HTML and XML have already been extended to support semantics within scientific domains, e.g. the Systems Biology Markup Language SBML (Hucka *et al.* 2003). The simplest semantic KR language for information exchange is the Re-

---

[2] Barry Smith has been an influential philosopher, producing a realist Aristotelian ontology. Nicholson and Dupré (2018) collect views on a process-based understanding of biology, including how this may impact bioontologies.

[3] Goble and Stevens (2008) report that John Quackenbush describes standards as "blue collar science" adding that "No-one will win a Nobel Prize for defining a workable format standard" (688). See also the compelling piece of Goble and Wroe (2004): This compares life scientists' and computer scientists' 'feud' to that between the Montagues and the Capulets in *Romeo and Juliet*, ; that is, as a feud blocking a great romance.

source Description Framework (RDF), which uses triples, of the form "(subject, predicate, object)", to represent information (Antezana *et al.* 2009, 397). Ontology languages can structure RDF further, and enable operations on them (cf. RDF Schema RDFS, or Web Ontology Language OWL).[4]

Scientific KM infrastructures are being developed through work upstream—by authors sharing "knowledge" through standardised formats, downstream—by curators and database managers who extract and store "knowledge" using KM tools, and midstream—by scientists sourcing and analysing "knowledge" from online resources. Imposing standards on scientific knowledge production aims to align second-order KM infrastructures with first-order knowledge production to "enhance"—make more precise, faster, larger-scale-the production of knowledge on the first-order (cf. recently Wilkinson *et al.* 2016; Edwards *et al.* 2007).

Why not just see KM as a tool for science, instead of a scientific research field itself? Scientific KM deserves the name 'science', as it promises to enable new first-order knowledge: it combines knowledge from information science and statistics with knowledge of a target epistemic domain's native epistemic standards to ensure that first-order knowledge and data are managed in ways that can ensure their validity, relevance and ethos. Certainly, second-order KM relies on first-order knowledge for its existence—there must be some kind of 'knowledge' there to manage. Yet scientific KM is developing science (biology, history, economics, …) on a meta-level, through codifying and managing first-order scientific activity explicitly and systematically. In this mode, KM-enabled science is like a snake biting its own tail, seeking to grow by re-sourcing its 'own' scientific activity in new, scientific, digitally mediated ways.

But is scientific growth possible this way? How can second-order KM add to first-order knowledge? We explore this question through a case and example.

2. *Scientific Knowledge Management producing first-order biological knowledge*

To illustrate the work of scientific KM and its impact we examine KM in the field of systems biology. As mentioned, our work is based on empirical, philosophical research.

Over the course of January 2012 to September 2013 Efstathiou participated daily in the work environment of Lægreid's lab, sharing office space with project members, attending and presenting in lab meetings, and following computational modeling work while also observing animal modeling in another lab (cf. Efstathiou 2018, 2019). Participation included observing people in their work environment and interviewing them, formally and informally, that is, using structured and unstructured interview formats. For this article we draw on six in-depth interviews pursued by Efstathiou, including with Lægreid and Kuiper and completed in the Spring and Fall of 2012, three of which are quoted here. Besides co-authors, one of our interviewees, 'Luke', has been a key informant, offering opportunities for several informal interviews in the period of the project.

---

[4] Such digital KM tools are also being developed for other fields, such as archaeology and the humanities, but our focus here is bioscience.

Our qualitative analysis of interview material focused on the use of the term "knowledge", and "knowing". We coded for different uses of this term, identifying manifest and operative concepts of knowledge in these domains (Haslanger 2005), i.e. definitions of knowledge in KM textbooks, and what conceptions of knowledge are "founded" and operative in getting KM work done (Haslanger 2005; Efstathiou 2012). Our focus was the practical wielding of the word "knowledge": How do KM researchers apply in practice, in bodily practice, the term "knowledge"? We put special attention on whether usage varied across disciplines.

In articulating a logic in the social practice of computational KM we claim that researchers operate with a sense of "knowledge" that locates it in the actual text of articles, as explicated facts and information, and further as arising from appropriately annotated data and metadata. This practice provides a way of working with knowledge as a thing, a resource to be extracted and organized from texts via the help of computers making new results possible. What our method does not support is a claim about what all or most KM researchers think. Rather, what we are communicating to the reader is one kind of social practice of knowledge and new conceptions of knowledge that emerge from this material, and which may apply elsewhere. Finally we choose to illustrate these ideas through our empirical material and organizational KM theory instead of starting with epistemological discussions in philosophy, as our focus are situated linguistic-embodied practices specific to KM.

We dub the group led by Lægreid the GAstrin BIology group, or GABI, and the group led by Kuiper as the SEmantic Systems Biology group, or SEB. The size of the groups is comparable, and has varied in this period between 5 to 10 members. GABI members are primarily trained in molecular biology, and lab-bench science. SEB members rely primarily on computational training, though several have a mixed background including biology. We use gendered acronyms to signal the gender balance in these groups. At the time of writing GABI is led by and includes a majority of scientists who identify as female, while SEB is led by and includes almost exclusively scientists who identify as male, with women in junior positions—profiles typical of respectively molecular biology and computer science work. Both groups include international members, though SEB is significantly more international. Using KM capabilities, GABI and SEB members are working to understand how mammalian cells respond to stimulation by the hormone gastrin. They are doing so within the frame of systems biology.

2.1. Systems biology and KM

Systems biology is a bioscience approach that has flourished in the paradigm of genomics (Powell *et al.* 2009; Keller 2005). The completion of the Human Genome Project in the early 2000s made clear that genes cannot account for biological complexity[5], and produced tools for sourcing more and more—omics data in need of accounting for (Blake and Bult 2006). From a field of experimental science purporting itself to be too complex to admit mathematical formalisation, biology is now arguably too complex not to try (Green 2017).

---

[5] A typical human-centric and gene-centric way to capture missing information is to compare the HGP findings on the genome of humans estimated at 30K genes with the weed Arabidopsis that counts 26K genes.

Systems biology factors into the study of biology some of the complexity, multi-layeredness and multi-causality that biological systems seem to have by combining molecular biology with methods from mathematics, physics and computer science. It negotiates contrasting commitments to abstraction among these epistemic communities (Keller 2002) to help understand biological systems as multi-composed and dynamic (Calvert 2010; Calvert and Fujimura 2011).

Computational tools are becoming crucial for the study of multi-component biological systems. The systems biology of a few components has been pursued for decades (Keller 2002; Peter and Davidson 2012) but approaches building on large-scale—omics data emerged only in the new century (Boogerd *et al.* 2007; Green and Wolkenhauer 2013). Computation is considered crucial for mathematical simulation and reasoning about large-scale systems, and for managing knowledge about hundreds of components at the same time.

But how is "knowledge" understood within KM-enabled biological practice? We cannot answer that question in general, but we consider some accounts by SEB researchers.

2.2. "Knowledge" in a semantic systems biology context

The review used earlier to map key tools in the field of scientific KM is co-authored by SEB members. Here is how the authors manifestly define knowledge:

> The concept of data came into prominence relatively recently, mainly due to the widespread use of the information and communication technologies (ICT) and the advent of modern empirical technologies that outpour huge amounts of data. Data should not be confused with **knowledge**—the former is just a collection of facts that require interpretation in order to be converted into knowledge. Thus, **knowledge is data plus an interpretation of its meaning** (Antezana *et al.* 2009, 392; emphasis added).

"Knowledge" is here juxtaposed to "data" and readers are warned against confusing the two. Knowledge can only be derived from an "interpretation" of the meaning of data. What does this involve? Not any interpretation goes!

We often need to specify the meaning of a word by attending to its use-context. If data are numbers or labels, knowledge is similarly described as possible to obtain by supplementing data with context.

> To give an example, consider the output of a microarray experiment. This is pure **data**, a matrix of labels and numbers that **conveys no meaning** to the human mind. A subsequent analysis of the data may reveal that a certain group of genes is over-expressed under certain conditions; if this finding would be based on experimental evidence obtained through accepted analysis approaches and have statistical significance, this would comply with the conditions above and constitute **a piece of knowledge**. Obviously, the same set of data may afford many alternative interpretations. Therefore, the concept of **'provenance'**, keeping track of how pieces of knowledge came to be, is crucial for KM. (Antezana *et al.* 2009, 392-393; emphasis added).

Providing context happens by specifying data provenance. This is an epistemologically thick concept as it is meant to keep track of the experimental analysis approaches used to derive the data. Data provenance is understood to provide evidence and thus to help

choose a valid data interpretation and to convert data into "pieces of knowledge"—note the metaphorical parsing of knowledge into bits. This in effect involves handling extra data about the data, or 'metadata', for instance when the data in question were obtained, by what experimental procedure, on what material:

> Numbers themselves [**data**] are meaningless, but knowing that the column with numbers depicts quantified fluorescence from a microarray experiment done on a breast tumour RNA extract allows one to **interpret** these as proxies for gene activity, if one also knows that each row represents one specific gene. (Comment on text, 12.10.14; emphasis added).

Knowledge is conceived as interpreted, or contextualised data (numbers or facts), where the contextualisation happens via the provision of metadata that help specify the provenance of these data to convert them into knowledge. How particular is this understanding of knowledge to SEB members?

In 2007 Chaim Zins published a Critical Delphi study of 150 information scientists specifically to analyze their definition of "three key concepts" (497): data, information and knowledge. Zins (2007) reports that in their majority responses conceived of "knowledge" as 'nonmetaphysical', i.e. as accessible to epistemic scrutiny, as 'cognitive-based', i.e. concerning states of mind, or meaning and intention, as 'propositional', i.e. as distinct from practical knowledge or knowledge by acquaintance, and last as 'human-centered', i.e. as pertaining to humans as opposed to other systems (487-488). The majority of respondents further agreed that data, information and knowledge are part of a continuum, where "data are the raw material for information, and information is the raw material for knowledge" (Zins 2007, 497; the existence of a Wikipedia entry on the "DIKW 'pyramid' of Data, Information, Knowledge and Wisdom" further indicates the typicality of this notion).

The manifest concept of knowledge defined among SEB members seems to agree with results in Zins (2007): knowledge is perceived as accessible to epistemic scrutiny, delivered by epistemic work, such as providing context to data, and as cognitive-based, "the interpretation of meaning", instead of by smelling, touching or being with data. Even if bioinformaticians tacitly know how to handle data, the definition of knowledge they work with is of it as a cognitive, intellectual output. However there is a point at which SEB members diverge from the results of Zins (2007).

SEB develop computational, semantic approaches to KM, using the developing Semantic Web. In this context, "knowledge" is not understood as human-centered, as Zins (2007) claims, but as accessible by and communicable among computers.

> Traditionally, the **interpretation** [of the meaning of data] was carried out by a human being; however, today the interpretation of large-scale data sets is typically **only possible with the help of computers** because of the sheer volume of data. ... KM is the process of systematically capturing, structuring, retaining and reusing information to develop an understanding of how a particular system (e.g. an organelle or a pathway) works, and subsequently to convey this information meaningfully to other information systems (**knowledge distribution**). (Antezana *et al.* 2009, 392, 393; emphasis added).

In this case, knowledge derived from large-scale data is described as "only possible with the help of computers", and further, as possible to "distribute" to other information systems.

Knowledge is thus not understood as "human-centered" but as possible and exchangeable, at times only, via computational means. This could be a matter of SEB's research focus, it may track changing perceptions in information science, or it might be that everyday and technical concepts of knowledge are not well kept apart in the inquiry.

2.3. Founded knowledge concepts

Ideas about knowledge appear here as "founded" in the epistemic practice of computational KM. Founded concepts are defined by Sophia Efstathiou as "transfigurations" of everyday ideas, following operations that gear them to work as technical, scientific concepts (2012, 2016). Founding a concept in a scientific domain happens through actions that can seem natural to practitioners, like (Efstathiou 2016, 53):

— focusing the concept on an ontological domain of interest
— expressing a concept in terms 'native' to a scientific domain
— operationalizing or devising ways to measure a concept
— discussing or publishing about this concept with colleagues.

Founding is "done" when the original idea is possible to find within the scientific domain as a scientific concept. Efstathiou calls the result "found science" by analogy to found art.

It appears that "knowledge" operates as a technical, founded concept in KM work: it re-articulates an everyday idea of knowledge to fit the epistemic cultures of computational science. To track how founding could happen here consider a manifest definition of an everyday idea of knowledge sourced from the Oxford English Dictionary[6]. Two main meanings of 'knowledge', ordinarily understood, are specified there:

— Facts, information and skills acquired through experience or education; the theoretical or practical **understanding** of a subject—e.g. I have good knowledge of grammar.
— Awareness or **familiarity** gained by **experience**—e.g. Sílvia's knowledge of human nature is remarkable—she can always read people.

Following this definition we can say that, manifestly, knowledge is ordinarily understood to involve learning facts, information or skills through education, or developing familiarity and awareness through personal experience.

We here propose that founding "knowledge" as a technical idea within KM happens by focusing on the ontological domain of facts and information, i.e. on knowledge as a phenomenon concerning the theoretical and practical understanding of a topic. This narrows the ontological scope of the everyday idea, to exclude informal, experiential and personal dimensions of knowledge. The concept becomes honed into those aspects of knowledge that are relevant for information science: knowledge then *is*, in this domain, facts and information. This specification allows a concept of knowledge to be further founded within computational KM by translating 'facts and information' in terms native to computer sci-

---

6   The definition is available at: http://oxforddictionaries.com/definition/english/knowledge

ence like 'data', 'metadata' and 'provenance', which allows the concept to be further operationalized via appropriate Knowledge Representations, ontologies and relevant syntax/semantics.

These two founded technical concepts: *explicated knowledge* (facts and information in the scientific literature) and *computable knowledge* (appropriately derived data and metadata) allow KM researchers to approach knowledge as (always already) a computational phenomenon. What power can founded concepts of knowledge afford working biologists? Consider a perspective from collaborators in SEB.

SEB member 'Ari' worked in conservation biology in India before his Masters in Bioinformatics (Interview, 3.10.12). Ari realised how important "handling data" is, while in the field. He worked with big fruit-eating bats, a specialist population feeding and living in only specific habitats, (like he is, he jokes, as a vegetarian in Scandinavia), and was also involved in a behavioural study of arachnids (spiders) "as big as my palm" (Interview, 3.10.12). Ari recalls that different research groups in the same research community often used different guidelines making it difficult for data from one institute to fit another's standards. He recalls how challenging it was to get from local to national data on the same species, especially to combine data from the North and South of India: "The North-South divide in India is sharp -in culture." (Interview, 3.10.12). Coming to his current work in semantic systems biology Ari explains:

> It is part of human nature: we are ambiguous in the way we say things. Semantic Web **connects data unambiguously and meaningfully**, with meaning attached to **context** so that they can be changed or agreed upon. (Interview, 3.10.12; emphasis added).

How can Semantic Web technologies help humans communicate, "unambiguously" and "meaningfully" here?

> The larger mission of Semantic Web is to convert stuff to **entity-based content**. So for example, when you say "Sophia", it should present YOU. "Sophia is—a person", "is—a biological entity", "is—a woman", "is-part-of-the Crossover Research group"; these would be different **relations** built into the knowledgebase to identify YOU. (Interview, 3.10.12; emphases added).

Changing data and agreeing on data, as biologists need to do, is to be mediated and facilitated by making data unambiguous and 'known' first for computers and networks of computers. The context where biological data would be given meaning is, in this case, a mixed biological and semantic web context, where "knowing" involves properly identifying things and relating them to other identified things, through identified relations.

This is a founded concept of knowledge as computable, from data plus appropriate metadata, which is markedly foreign to biological practice. The concept seems possible to smuggle into first-order biological practice, through a prior founding of 'knowledge' as facts and information explicated in published texts.

The next section illustrates how founded technical concepts of knowledge as *explicated* and *computable* facilitate KM practices: they aid KM in deriving new first-order biological knowledge, in new ways.

2.4. KM impacting first-order biological knowledge: Assembling "knowledge" into a model

Consider a central question in GABI research:

— What happens to a cell when it is stimulated by the hormone gastrin?

Gastrin is released in the gastric mucosa, the lining of digestive organs, and it contributes to physiological processes like digestion, appetite control and body weight regulation. It is also associated with several diseases including cancer. GABI researchers are interested in how the CholeCystoKinin 2 cellular Receptor (referred to as "CCKR" among researchers) mediates these responses from inside the body to cell nuclei and genomes. GABI have pursued wet-lab and increasingly KM-based research to answer their central question.
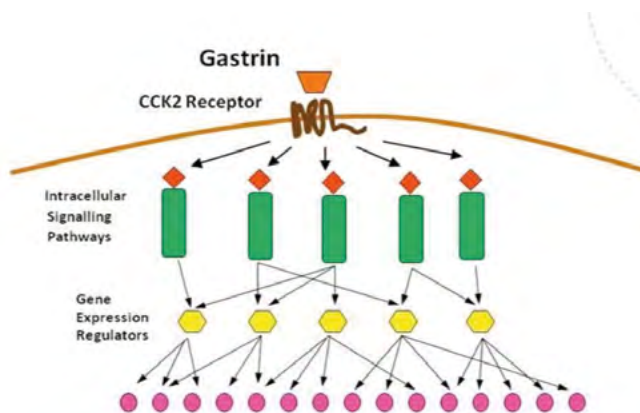

Figure 1

Drawing developed by GABI members to represent gastrin-mediated signalling and regulation of gene expression. The hormone gastrin interacts with its specific CCK2 receptor, which transduces the gastrin signal through the cell membrane (curved line), and via signalling pathways (diamonds on rectangles) and gene expression regulators (hexagons) down to gene activities (ovals). [Pulled from Lægreid's presentation slides.]

To better understand what happens inside mammalian cells stimulated by gastrin (Figure 1), the molecular biologist and GABI member 'Luke' collected "published knowledge" about all cellular components (genes, proteins, RNAs, metabolites) described to respond to gastrin in different experimental systems (different mammalian cell lines, from different organisms, at different conditions) (field notes, September 2012). Luke created a "knowledge assembly" model, operating with assumptions about the extractability and compose-ability of biological "knowledge", across experimental contexts (Tripathi *et al.* 2015). In publications the model is referred to as a "signalling network" and "signalling map" primarily, instead of a "knowledge assembly" model, which was how the model was described in conversation. The epithet "knowledge-assembly" makes clear the second-order application of KM tools in building the model. Calling the model a "signalling network" or "map" points instead to the first-order biological target under model representation: cellular signalling processes.
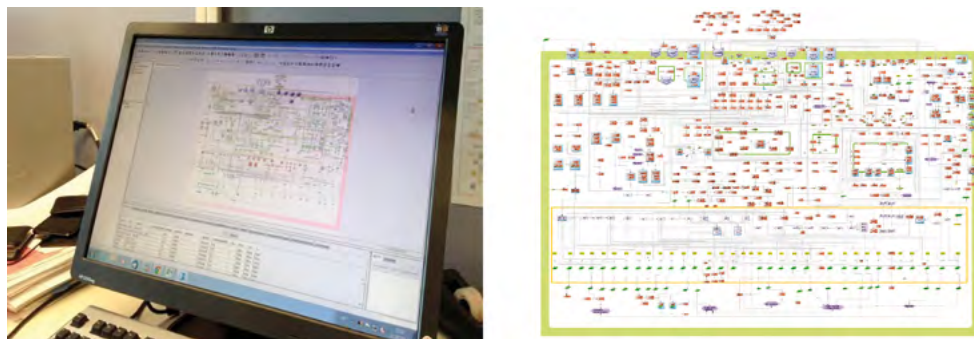
Figure 2

CellDesigner model on screen (left), and in print (right). The model encompasses a total of 530 proteins and genes (various shapes) linked by 413 interactions (lines). The entity names are hyperlinked to standard bioontologies and databases, and causal regulatory information is connected to PubMed IDs of the scientific articles from which the information was collected (Tripathi *et al.* 2015).

Central to representing this "knowledge" is the pathway-editing software CellDesigner™ (Funahashi *et al.* 2003; Kitano 2003) (Figure 2). CellDesigner was created in Hiroaki Kitano's laboratory at Tokyo University (Available online at http://celldesigner. org/). Kitano is one of the people leading the current computationally heavy and semantically integrated vision of systems biology (cf. Kitano 2002). Why should biologists use this tool? Kitano expresses the need to have computationally 'structured' visual representations for molecular, gene or protein networks and interactions as follows (2003):

> Currently **knowledge** on molecular interactions is mostly described either by written text or by traditional cartoon-like diagrams. Written text is inherently ambiguous, and results have had to be re-interpreted by each reader of the article. Most authors of biological papers use arrow-headed lines to indicate activation and inhibition, respectively, with mixed and often inconsistent semantics. However, traditional diagrams are informal, often confusing, and much information is lost. Thus the urgent task is to provide a set of notations that have **powerful expression capability** and are highly **readable** for biochemical and gene regulatory networks (169, emphasis added).

Kitano's argument echoes Ari's remarks: standard biological communication through text and diagrams is "ambiguous". How is published knowledge "disambiguated" by CellDesigner? By providing standardised formats for its representation and by thereby fixing rules for its interpretation. The shapes, or "glyphs" used by CellDesigner are generally accepted as a standard for the visual representation of biological networks, known as Systems Biology Graphical Notation (SBGN—Le Novère *et al.* 2009). CellDesigner enables a computational simulation of biological 'knowledge', understood here as facts and information described by written text and diagrams in the literature. CellDesigner uses the KR language generally accepted for such simulations, Systems Biology Markup Language (SBML—Hucka *et al.* 2003). The representational choices offered by CellDesigner look similar to how biologists would "anyway" draw diagrams, yet CellDesigner enables the computational comparison, compilation and sharing of

these models, and the further interpretation of 'knowledge' explicated and made computable in them.

CellDesigner helps manage textually *explicated* knowledge by hosting it in computationally manageable, standardised, *computable* formats. But computational tools are also needed to feed knowledge into the model. Luke searched scientific publications for combinations of the hormones "cholecystokinin (CCK)" and its receptor "CCK1R" and of "gastrin (G-17)" and its receptor "CCK2R", through PubMed and various literature-mining tools, e.g. LitInspector and iHOP (Tripathi *et al.* 2015, 2). First-order biological knowledge developed by training and practical experience with standard wet-lab work is crucial for adequately curating data resources and for model building. More than 250 of circa 1200 articles were selected as useful references, by Luke, because they contained what was deemed, by curator judgment call, as "good evidence" that the reported signalling event is mediated by the interaction of gastrin with its receptor, and provided sufficient signalling information allowing for linkage of a new model component to its upstream and/or downstream regulators and effectors (Tripathi *et al.* 2015, 2).

Not any knowledge claim explicated in scientific text will do. References selected here were "extracted" by a team of five trained biologists from GABI and SEB who read the literature, and represented the information and facts explicated in this literature appropriately via the CellDesigner platform. The five team members individually read and ranked claims in the final selection in terms of their confidence in these claims as "OK, DISCUSSION, INCORRECT", and they further critically discussed how to represent reactions, components and cellular localisations through the software (cf. Tripathi *et al.* 2015, 3). This scale of parallel curation is rather uncommon in large-scale biocuration, given how limited current resources for biocurators are. The protocol followed here is thus atypically rigorous and very much reliant on the biological expertise of the curators in adequately translating between explicated and computable knowledge.
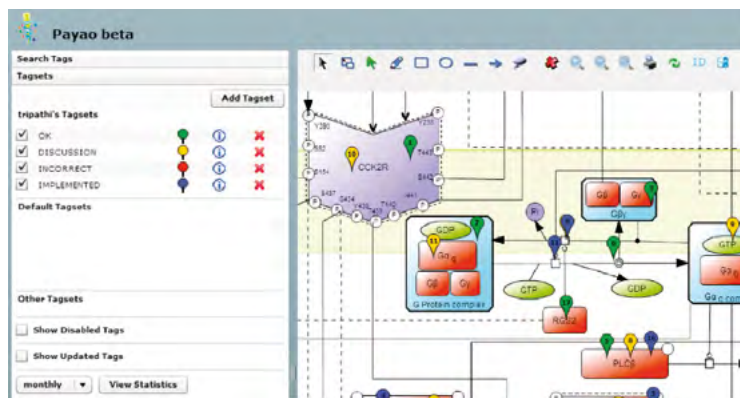

Figure 3

Collaborative model-construction on the PAYAO platform. On the left hand panel we see a coding of map "points" tagging model components that group members discussed: the tagsets used are 'OK' (green), 'DISCUSSION' (yellow), 'INCORRECT' (red) and 'IMPLEMENTED' (blue). (Reproduced with permission, Tripathi *et al.* 2015, figure 1)

2.5. KM as epistemically productive and practice-dependent

Sabina Leonelli (2014) argues that the prospects of fully automating and replacing the capacities of scientists to assess and interpret data are highly doubtful but that computational tools facilitate collaborative thinking among working teams of scientists (399-400). Collaborative model-construction by GABI and SEB members was indeed a crucial outcome of using the community-curation platform associated with CellDesigner, PAYAO (Figure 3). Still, and though we agree with Leonelli that full automation is highly unlikely, these computational tools are not epistemically inert.

Miles MacLeod and Nancy Nersessian have analysed building dynamic network models within Integrative Systems Biology as "modelling from the ground up" (2013). The model they focus on is similar to the CCKR model but with added work by engineers to model these interactions dynamically. In their analysis, this type of model-building involves approximating the causal structure of a phenomenon by assembling existing information about its components—as opposed to generating the phenomenon from simpler theoretical rules. This approach can be theory "light", following pragmatic constraints (see also Leonelli *et al.* 2012). But in the case of MacLeod and Nersessian (2013), constructing such models involved engineers with no biological knowledge performing similar literature searches as Luke did in our case. For example, describing the construction of dynamic models of such pathways by an engineer, MacLeod and Nersessian (2013) say: "In each case, the pathway given to her by her collaborators was insufficient given her modelling goals, and she was forced to pull in whatever pieces of information she could find from literature searches and databases about similar systems or the molecules involved, in order to generate a pathway that mapped the dominant dynamic elements" (541). Lacking an adequate knowledge about biology could mean that when selecting what references to include in a pathway model, all one can rely on is KM resources.

Further scientific inferences were made 'automatically' also in our case study. Once explicated biological knowledge was curated and represented in CellDesigner, the map was analysed using computational tools, in this case Cytoscape and the BiNoM plugin (Shannon *et al.* 2003). Decomposing the map into sub-networks using a "pruning" software function "revealed" 18 modules that were "higher-level structures" of the signalling map (Tripathi *et al.* 2015). The software helped to analyse what happens in a cell stimulated by gastrin by isolating different signalling pathways linked to particular outcomes, like proliferation, migration and apoptosis. And still further data, besides the explicated literature-curated and computationally analysed knowledge, was brought in to explore these interactions.

Large-scale Protein-Protein Interaction (PPI) data was downloaded from databases using the webservice Proteomics Standard Initiative Common QUery InterfaCe, PSIC-QUIC. The selection was filtered using controlled vocabulary terms to just include binary physical interactions, and the data was added to the literature-based map, to enable the further biological interpretation of the interactions represented there. Combining interaction data with topological network analysis, and using their biological expertise, GABI and SEB researchers identified seventy proteins, which "represent experimentally testable hypotheses for gaining new knowledge on gastrin- and cholecystokinin receptor signalling" (Tripathi *et al.* 2015, 1). Seventy proteins may seem like a lot of proteins to ask biologists to run

individual experiments on, but in a field seeking to explore thousands of biological interactions it is a small number.

In sum, computer-based scientific KM enables sourcing, representing and analysing the biological literature and it informs hypotheses to test in the lab. Visual communication and representation practices are key both for information sharing and for building communal vision, especially in multi-disciplinary teams (Carusi 2011; Coopmans *et al.* 2014). And they are epistemologically productive. A biologist may be capable of mentally picturing molecular interactions in small-scale models, but this is challenging for large-scale models. Processing and depicting biological knowledge about molecular interactions through CellDesigner or Cytoscape transforms practices of network construction and analysis in biology. It develops the know-how of biologists as users of these tools, while transforming what was originally sourced as first-order knowledge explicated in the literature, into data of computational value, for the purposes of assembling and analysing this knowledge from a higher-level in a way that can feed it back into biological inquiry.

This type of work is seen as especially crucial for the field of systems biology. The result is computationally accessible data (the model itself) and further explicated knowledge (the accompanying article). Note that the epistemic validity of KM-enabled systems biology still depends on experimental knowledge of biology: this informs first, creating KM infrastructures through adequately aligning the standards, languages and structures required by computational tools with what gives meaning to working biologists, and second developing epistemically adequate protocols for using KM tools within biological research.

3. *Overcoming limitations of KM knowledge concepts and epistemologies*

Second-order scientific KM is transforming first-order biological knowledge practices. But there is a cost to these tools, we caution. The concepts enabling KM researchers to think of knowledge as possible to source, "extract" from the literature and to "assemble" and "distribute" in computational, semi-automated ways prime an understanding of knowledge as an objective object. Computational KM thus risks losing track of the context-sensitivity and contestability of scientific knowing unless practice-based biological knowledge is openly appreciated as intrinsic to the validity and validation of these tools.

3.1. Scientific KM and collaborative labour

How will 21st century biology make appropriate use of 'its' knowledge? KM work mixes biological and computational expertise, at different levels of visibility and importance. Manifestly biological knowledge is the key epistemic resource offered by network models. Yet this knowledge is already processed in computational formats: 'marked' and 'marked-up' as computer comprehensible.

Kitano assumes that biological knowledge formatted in CellDesigner is possible to comprehend by biologists. But pointing to a space in a "knowledge assembly" model is not by default meaningful to a molecular biologist—at least not when compared to experimental observation. When asked about people's responses to the CCKR model, Luke answers that people are sometimes "amused" (Interview, 31.5.12). Sometimes they find the model

"scary": as "the very simplified" version of the pathways is the usual picture they have (Interview, 31.5.12). Luke adds:

> Everyone knows that cell machinery is very complicated, that wiring inside the cell is very complicated. So people [molecular biologists] want to focus on their own domain [and say]: "If I'm working on this component why care about the rest?" (Interview, 31.5.12).

The value of computationally founded knowledge of a cellular space is contrasted with the value of knowing components one is familiar with, experimentally. The vision of systems biology is that of knowing a whole system. But perhaps a "cellular signalling map" is frightening for a molecular biologist who does not feel lost (or who is happy to work with tunnel vision)!

GABI member 'Silja' was trained in mathematics and computer science but switched to biology and biochemistry. Silja worked with mathematicians in the early days of microarray experiments to distinguish signal from noise. She recalls the real need for such tools, emphasising the risk of making computer scientists' labour invisible in biology.

> At first we were asking them to work for us, but then we had a project together.... I keep saying: "If you need someone to work for you, you need an engineer". But it is not possible to collaborate and keep asking them [bioinformaticians] to work for you—we cannot always be leading. They can be main authors, supervisors. (Interview, 10.10.12).

Silja was involved in extensive microarray time-series experiments, producing temporal data coveted by both experimentalists and computational biologists. GABI research with this data has shown that gastrin upregulates genes that may be involved in different physiological processes, including tumorigenesis, proliferation, endoplasmic reticulum stress, anti-apoptosis, differentiation and migration. In our conversation Silja shared her future plans to use the data to further explore protein expression and cell fates in *in vitro* and *in vivo* models.

Why not use the time-series results *in silico*, to develop KM tools? Silja reports that she was invited to reanalyze the data and "get more **knowledge**" together with SEB researchers. She adds:

> But I am more interested in using the data. That is why I am now working with 'Tanja' and 'Hannah' [biologists], trying to **understand** the data more... I like experimental (wet-lab) work as well, and I am not so eager about spending considerable more time on generating bioinformatics tools. (Interview, 10.10.12; emphasis added).

Silja juxtaposes "understanding the data" with using the data to get more "knowledge". The term "knowledge" here specifies an outcome of computational processing, indicating that the founded concept is operating in the lab and also in the work of biologists. This "knowledge" is contrasted in the next sentence with what, in Silja's view, offers an "understanding" of the data: "using" the data to do further experimental—wet lab—work. This could indicate a contrast between the founded knowledge that results from computational work with (really) understanding the data, through experimental molecular biology. And note also the shift in the labour dynamics here: at this moment in time, a biologist could also feel

that the ownership of her labour is at stake, as computational biologists are 'using' biological data.[7]

It need not be that Silja is critical to KM development; simply the joy and familiarity that experimental work provide may be what drive experimental biologists to continue their work. But it certainly seems that practices and values are not smoothly shared across computational and experimental domains, posing a choice: How can biology best manage its knowledge? Is computational KM enhancing or compromising traditional, first-order knowledge production?

We reflect on these questions in the next section. Our suggestion is that KM may enhance first-order knowledge production if it embraces and acknowledges that practice-based epistemological approaches are part of its practice.

3.2. Organisational KM: Towards a practice-based epistemology for scientific KM

Donald Hislop's account of organisational KM distinguishes two theories of knowledge (2013, Chapters 2 and 3). "Objectivist" epistemologies consider knowledge as an object: some thing that can be separated from the knowers, codified, stored and trafficked, objectively. "Practice-based" epistemologies instead consider knowledge as embedded in and inseparable from people's practices, bodies and cultures and as intrinsically social and negotiated. This overlaps with philosophical distinctions between 'explicit' and 'tacit' knowledge (Polanyi 1967), and between propositional and non-propositional or embodied knowledge, what Gilbert Ryle called "knowing that" versus "knowing how" (Ryle 1949). The importance of practice-based knowledge is highlighted by history and philosophy of biology—most notably in Keller's discussion of Barbara McClintock's "feeling for" her corn plants (Keller 1983), but also specifically in the context of biocuration (Leonelli 2014). Here we are interested in situating this distinction instead as a part of the theoretical tradition of organisational KM which is closer to our informants' work practice.

Life science KM, at its word, seems to imply an objectivist epistemology. According to Hislop (2013) objectivist epistemologies assume/enforce four claims: 1. knowledge is an *object*, 2. knowledge is *objective*, 3. *explicit* knowledge is better than tacit knowledge, 4. knowledge is *cognitive* (18-19). SEB members and their GABI partners involved in our study refer to knowledge as a thing, considered possible to separate from those who have it, to "extract", codify and analyse it. Semantic web tools seem to promise 'objectivity' as knowledge is to be "disambiguated", and thus possible to share among scientists beyond particular (idiosyncratic, subjective) terminologies, national/cultural contexts or work cultures. Assembling and representing explicated knowledge is seen as 'presenting the facts' and thus knowledge-assembly models can become synonymous to "maps" of actual cellular spaces. There is no doubt that biological knowledge and computer science knowledge can be tacit and that both are crucial for epistemically adequate KM. KM sci-

---

[7] Different cultures of ownership among knowledge managers/informaticians and experimentalists are discussed by Bruno Strasser, e.g. 2011.

entists would not deny this. Yet another type of knowledge takes the spotlight as valuable here.

Efforts are put into further 'automating' quality assessments, explicating and codifying practice-based knowledge via "evidence codes", and other metadata, for biology experts to be able to "interpret" data into knowledge faster, using computational tools to help reason to an outcome which is considered cognitive as opposed to embodied. Overall, and despite the importance of experimentation, KM purports to be able to manage experimentally produced knowledge, "better".

What would KM look like instead from the perspective of a "practice-based epistemology"? Would a practice-based epistemology even be possible, given how KM tools have been developed? Practice-based epistemology highlights aspects of knowledge and knowing that are tacit and embodied, and that cohere with the values of feminist epistemology (e.g. Anderson 1995). In this view: 1. knowledge is a *process*, 2. Explicating knowledge is *incomplete*, 3. Knowledge is *multidimensional*, 4. Knowledge is *socially* produced, *uncertain* and *political* (Hislop 2013, 32-41). From this epistemological perspective, there can be many frames for understanding biological knowledge. First, biological knowledge could show up as embedded in biological practices, occurring in on-going human-non-human laboratory activities whereby knowing and doing are hard to dichotomise, and where objects and classifications are made and remade depending on the interests at hand (cf. Dupré 1993). In this approach, KM tool creation would need to be seen as intrinsically revise-able, and durational, if not using process-based ontologies. Further, a practice-based epistemology challenges the assumption that biological knowledge can be fully explicated and codified, implying that knowledge possible to manage via current computational KM tools would be by default incomplete. Following a practice-based epistemology, developing KM tools involves inherently ambiguity, uncertainty, and the exercise of judgement on the part of those pursuing knowledge –professionals as well as the technologies they relegate decisions to. Third, in this view, knowledge is *multidimensional* both embodied and intellectual, tacit and explicit, collective and individual, developing and static. Managing to 'know' biology within biological institutions would need to recognise the multiple expressions, "ambiguity", and inconsistencies, also as part of getting better knowledge. Fourth, a practice-based understanding of knowledge views it as *socially constituted*, pursued in communities and varying across disciplinary and national cultures for legitimate, indeed unavoidable, reasons. National and cultural factors impact how biological knowledge is developed, on what topics, for how much funding, with what expectations, on whose bodies. In this frame, knowledge is visible as *political*, meaning that differentiations between knowing and not knowing groups or people, humans and nonhumans, come with polarisation, inequalities, conflict and negotiations of power.

As already stated, our material indicates that knowledge practices within current computational KM rely on objectivist epistemologies: understanding knowledge as cognitive, and objective and of added value when explicated. But perhaps KM need not operate with this view. The work of experimentalists, and biocurators to produce KM knowledge structures is very much embodied and situated and intrinsic to the quality assurance of KM utilisation protocols. In our case, Luke's and his four collaborators' labour to read and rank literature claims was intrinsic in sourcing "well-evidenced" "knowledge' to be further, semi-automatically, managed. Computational KM practices could openly appreciate themselves as part of an ecology of knowing that intrinsically involves practice-based, biological

knowing and experimentation in its uncertainty, corporeality and context. In this frame, collaborations between experimental and computational biologists would become an essential lifeline and quality assurer for KM, which could in return help manage knowledge better (Figure 4).

Figure 4

Drawing developed by SEB members to represent the semantic systems biology work-cycle (Left). Semantic systems biology operates on knowledge extracted from literature and databases, processing it computationally to develop new hypotheses that can be tested in biological experimental practice (Left). Our analysis here flags the 'yin' in the 'yang', and 'yang' in the 'yin' for this work to be properly balanced (right): practice-based knowledge is needed to support computational conclusions -- theoretical work is also operative in facilitating experimental work.
[Reproduced from Kuiper's presentation slides; see also figure 2 Antezana *et al.* 2009, 401.]

Conclusion

We have argued for one main point in this article. Computationally enabled knowledge management practices offer second-order scientific ways to derive new, first-order biological knowledge. We specified two founded concepts of knowledge enabling this work: a. knowledge conceived as *facts and information **explicated** in published scientific texts*, and b. knowledge conceived as ***computable** via appropriately derived data and metadata*. KM practices help transform biological knowledge into explicated knowledge with computational value, for instance structured as "signalling networks" that enable novel clustering and other graph analysis operations. This knowledge, though manageable, seems remote from traditional experimental knowing, but it should not. Experimental expertise, practice-based knowing though processual, uncertain, embodied and contestable are intrinsic to securing the validity of manageable knowledge as knowledge.

Jim Grey, researcher and software designer in IBM and Microsoft, infamously heralded a new, "fourth paradigm" for scientific research: following theory-based, experiment-based and computation-based science we were entering an informatics-based science—a simplistic but powerful statement (Hey *et al.* 2009, xviii). Karl Popper, a man of clear physicalist and materialist persuasion also considered the move from 'subjective' knowledge to published theories in libraries as an evolutionary step in human development (1972). Yet he argued that the growth of knowledge must be in principle unpredictable: If one could predict how knowledge would grow and obtain the knowledge of tomorrow today, there would be no more growth to it (Popper 1972, 296-300). Perhaps then, KM visions such as those that Jim Gray pose for 21st century knowledge can be seen in these terms: automating scientific knowledge discovery were it to be possible would run the risk of killing—or at least stunting the growth—of knowledge.

To finish with the poetry of a.rawlings (2006, 42):

specify comma, question mark? dissect comma? intersect question mark, comma?

> *Collect, sort and frame text.*
> *How does a text fall asleep?*
> *Pinch meaning between morpheme and phoneme.*
> *How does text eat itself?*
> *Slide meaning into envelope; store in box with semanticide.*

comma, question mark specimen? comma dissection? question mark, comma cross-section?

REFERENCES

Anderson, Elizabeth. 1995. Knowledge, human interests, and objectivity in feminist epistemology. *Philosophical Topics* 23/2: 27-58.

Antezana, Erick, Martin Kuiper and Vladimir Mironov. 2009. Biological knowledge management: The emerging role of the semantic web technologies. *Briefings in bioinformatics* 10/4: 392-407.

Balmer, Andrew S., Jane Calvert, Claire Marris, Susan Molyneux-Hodgson, Emma Frow, Matthew Kearnes, Kate Bulpin, Pablo Schyfter, Adrian Mackenzie, Paul Martin. 2015. Taking roles in interdisciplinary collaborations: Reflections on working in post-ELSI spaces in the UK synthetic biology community. *Science & Technology Studies* 28/3: 3-25.

Blake, Judy and Carol Bult. 2006. Beyond the data deluge: Data integration and bio-ontologies. *Journal of Biomedical Informatics* 39: 314-320.

Boogerd, Fred C., Frank J. Bruggeman, Jan-Hendrik S. Hofmeyr, Hans V. Westerhoff. eds. 2007. *Systems biology: Philosophical foundations*. Amsterdam: Elsevier.

Calvert, Jane. 2010. Systems biology, interdisciplinarity and disciplinary identity. In: J.N. Parker, N. Vermeulen and B. Penders, eds. *Collaboration in the new life sciences,* 201-218. Aldershot, UK: Ashgate.

Calvert, Jane and Joan H. Fujimura. 2011. Calculating life? Duelling discourses in interdisciplinary systems biology. *Studies in History and Philosophy of Biological and Biomedical Sciences* 42/2: 155-163.

Carusi, Annamaria. 2011. Computational biology and the limits of shared vision. *Perspectives on Science* 19/3: 300-336.

Chalmers, David. 2009. Ontological anti-realism. In David Chalmers, David Manley, and Ryan Wasserman. eds. *Metametaphysics: New essays on the foundations of ontology*, 77-129. Oxford: Oxford University Press.

Cleveland, William S. 2001. Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review*. 69/1: 21-26.

Coopmans, Catelijne, Janet Vertesi , Michael Lynch & Steven Woolgar. eds. 2014. *Representation in scientific practice revisited*. MIT Press.

Peter, Isabelle S. and Eric H. Davidson. 2012. Transcriptional network logic: The systems biology of development. In Walhout Marian A.J., Marc Vidal and Job Dekker. eds. *Handbook of systems biology*, 211-228. Dordrecht: Elsevier.

Dupré, John. 1993. *The disorder of things: Metaphysical foundations of the disunity of science*. Cambridge and London: Harvard University Press.

Efstathiou, Sophia. 2012. How ordinary race concepts get to be usable in biomedical science: An account of founded race concepts. *Philosophy of Science* 79: 701-713.

Efstathiou, Sophia. 2016. Is it possible to give scientific solutions to Grand Challenges? On the idea of grand challenges for life science research. *Studies in History and Philosophy of Biological and Biomedical Sciences* 56: 48-61.

Efstathiou, Sophia. 2018. Im Angesicht der Gesichter: 'Technologien des Gesichtsverlusts' in der Tierforschung. In Wunsch Matthias, Martin Böhnert and Kristian Köchy. eds. *Philosophie der Tierforschung Volume 3: Milieus und Akteure*, 375-419. Freiburg und München: Verlag Karl Alber.

Efstathiou, Sophia. 2019. Facing animal research. Levinas and technologies of effacement. In Peter Atterton and Tamra Wright, eds. *Face to face with animals. Levinas and the animal question,* 139-164. Albany: SUNY Press

Edwards, Paul N., Steven Jackson, Goeffrey C. Bowker and Cory P. Knobel. 2007. *Understanding infrastructure: Dynamics, tensions, and design*. Ann Arbor: Deep Blue.

Funahashi, Akira, *et al.* 2003. CellDesigner: A process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 1: 159-162.

Gabrielsen, Ane Møller. 2018. Biocurators and reconfiguration of rrust in data-centric biology. Presentation. *Society for New and Emerging Technologies*, Maastricht: 25-27 June, 2018.

García-Sancho, Miguel. 2012. From the genetic to the computer program: the historicity of 'data' and 'computation' in the investigations on the nematode worm C. elegans (1963-1998). *Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1): 16-28.

Goble, Carol and Chris Wroe. 2004. The Montagues and the Capulets. *Comparative and Functional Genomics* 5: 623-632

Goble, Carol and Robert Stevens, (2008), State of the nation in data integration for bioinformatics, *Journal of Biomedical Informatics,* 41: 687-693.

Green, Sara. 2017. Introduction to philosophy of systems biology. In Green, Sara. Ed. *Philosophy of systems biology: Perspectives from scientists and philosophers*, 1-24. Dordrecht: Springer.

Green, Sara and Olaf Wolkenhauer. 2013. Tracing organizing principles: Learning from the history of systems biology. *History and Philosophy of the Life Sciences*. 35(4): 553-576.

Hackett, Elizabeth and Sally Haslanger. 2006. *Theorising feminisms: A reader*. New York: Oxford University Press.

Haslanger, Sally. 2005. What are we talking about? The semantics and politics of social kinds. *Hypatia* 20/4: 10-26.

Hey, Tony, Stewart Tansley and Kristine Tolle. 2009. *The fourth paradigm. Data-intensive scientific discovery*. Redmond, Washington: Microsoft Research. Available at: http://research.microsoft.com/en-us/collaboration/fourthparadigm

Hislop, Donald. 2013. *Knowledge management in organizations: A critical introduction*. Oxford: Oxford University Press.

Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., & the SBML Forum. 2003. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19/4: 524-531.

Keller, Evelyn Fox. 1983. *A Feeling for the organism*. New York: W.H. Freeman.

Keller, Evelyn Fox. 2002. *Making sense of life: Explaining biological development with models, metaphors, and machines*. Cambridge, MA: Harvard University Press.

Keller, Evelyn Fox. 2005. The Century Beyond the Gene. *Journal of Biosciences* 30/1: 3-10.

Kitano, Hiroaki. 2002. Systems biology: A brief overview. *Science* 295: 1662-1664.

Kitano, Hiroaki. 2003. A graphical notation for biochemical networks. *Biosilico* 1/5: 169-176.

Larsen, Peder Olesen and Markus von Ins. 2010. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics* 84: 575-603

Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M., Wimalaratne, S., Bergman, F.T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villeger, A., Boyd, S.E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T.C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D.B., Sander, C., Sauro, H., Snoep, J.L., Kohn, K., & Kitano, H. 2009. The systems biology graphical notation. *Nature Biotechnology* 27/8: 735-741.

Leonelli, Sabina. 2010. Documenting the emergence of bio-ontologies: or, why researching bioinformatics requires HPSSB. *History and Philosophy of the Life Sciences* 32/1: 105-126.

Leonelli, Sabina. 2014. Data interpretation in the digital age. *Perspectives on Science* 22/3: 397-417.

Leonelli, Sabina. 2016. *Data-centric biology: A philosophical study*. Chicago: University of Chicago Press.

Leonelli, Sabina. ed. 2012. Data-driven research in the biological and biomedical sciences. Special section in *Studies in History and Philosophy of Biological and Biomedical Sciences* 43/1: 1-316.

Lord, Phillip and Robert Stevens. 2010. Adding a little reality to building ontologies for biology. *PLoS ONE* 5(9): e12258

MacLeod, Miles and Nancy Nersessian. 2013. Building simulations from the ground up: modeling and theory in systems biology. *Philosophy of Science* 80: 533-556.

Nicholson, Daniel J. and John Dupré. 2018. *Everything flows: Towards a processual philosophy of biology.* Oxford: Oxford University Press.

Nydal, Rune, Sophia Efstathiou and Astrid Lægreid. 2012. Crossover research: Exploring a collaborative mode of integration. In Van Lente, Harro, Christopher Coenen, Torsten Fleischer, Kornelia Konrad, Lotte Krabbenborg, Colin Milburn, Francois Thoreau and Torben Z. Zülsdorf. Eds. *Little by little: Expansions of nanoscience and emerging technologies*, 181-194. Heidelberg: AKA Verlag.

Polanyi, Michael. 1967. *The Tacit dimension*. London: Routledge.

Popper, Karl. 1972. *Objective knowledge: An evolutionary approach*. Oxford: Oxford University Press.

Powel, Alexander, Maureen A. O'Malley, Staffan Müller-Wille, Jane Calvert and John Dupré. 2009. Disciplinary baptisms: A comparison of naming stories of genetics, molecular biology, genomics and systems biology. *History and Philosophy of the Life Sciences* 29/1: 5-32.

Price, Derek. J. de Solla. 1961. *Science since Babylon*. New Haven, Connecticut: Yale University Press.

Price, Derek. J. de Solla. 1963. *Little science. Big science.* New York: Columbia University Press.

rawlings, angela. 2006. *Wide slumber for lepidopterists*. Coach House Books.

Ryle, Gilbert. 1949. *The concept of mind*. Chicago, Illinois: The Chicago University Press.

Shannon, P. Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schiwikoski, B., Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13/11: 2498-504.

SINTEF. 2013. Big Data, for better or worse: 90% of world's data generated over last two years. *Science Daily*. 22 May 2013, Available at: <www.sciencedaily.com/releases/2013/05/130522085217.htm.

Stegmaier, Peter. 2009. The rock 'n' roll of knowledge co-production. *EMBO Reports* 10(2): 114-119.

Strasser, Bruno J. 2011. The experimenter's museum: Genbank, natural history, and the moral economies of biomedicine. *Isis* 102:60-96.

Tripathi, Sushil, Flobak Åsmund, Chawla Konika, Baudot Anaïs, Bruland Torunn, Thommesen Liv, Kuiper Martin and Astrid Lægreid. 2015. The gastrin and cholecystokinin receptors mediated signaling network: a scaffold for data analysis and new hypotheses on regulatory mechanisms, *BMC Systems Biology*, 9:40.

Van der Burg, Simone and Tsjalling Swierstra. Eds. 2013. *Ethics on the laboratory floor*. Palgrave: Macmillan.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Phillip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumen, Scott Edmunds, Chris T. Evelo, Righard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Pilippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Kathrine Wolstencroft, Jun Zhao and Barend Mons. 2016. Comment: The FAIR guiding principles for scientific data management and stewardship. *Nature: Scientific Data* 3:160018.
Zins, Chaim. 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology* 58(4): 479-493.

Acknowledgements

Sophia Efstathiou is Postdoctoral Fellow in the Programme for Applied Ethics, Institute for Philosophy and Religious Studies, NTNU. She works in philosophy of science, ethics of technology and art-based approaches to response-able innovation. Her latest publications include "Appreciation through use: How industrial technology articulates an 'ecology of values' around Norwegian seaweed", *Philosophy & Technology* (2018) (with B.K. Myskja), and "Facing Animal Research: Levinas and *technologies of effacement*", in Atterton P. & Wright T. (eds.) *Face-to-Face with Animals: Levinas and the Animal Question*. Albany: SUNY Press. Her work has received NSF, Max Planck and White fellowships.

Address: Postdoctoral Fellow, Institute of Philosophy and Religious Studies. Låven, Dragvoll, Norwegian University of Science and Technology/ Norges Teknisk-Naturvitenskapelige Universitet (NTNU), 7049 Trondheim, Norway. Email: sophia.efstathiou@ntnu.no.

Rune Nydal is Associate Professor at the Programme for Applied Ethics, Institute of Philosophy and Religious Studies, NTNU. He is a philosopher of science and technology, interested in the justifications and methods of interdisciplinary collaboration. Nydal teaches research ethics and the ethics of science and technology. He is a member of the Norwegian National Committee for Research Ethics in Science and Technology (NENT).

Address: Associate Professor, Institute of Philosophy and Religious Studies. Låven, Dragvoll, Norwegian University of Science and Technology/ Norges Teknisk-Naturvitenskapelige Universitet (NTNU), 7049 Trondheim, Norway. Email: rune.nydal@ntnu.no.

Astrid Lægreid is Professor at the Department of Clinical and Molecular Medicine, NTNU. She is a molecular biologist who pursues epistemic and socio-ethical questions as a central part of her efforts to transform her cancer research strategies from traditional molecular cell biology into a systems biology approach. Lægreid teaches systems biology to students across all disciplines.

Address: Professor, Institute of Molecular Medicine and Cancer Research. Norwegian University of Science and Technology/ Norges Teknisk-Naturvitenskapelige Universitet (NTNU), 7049 Trondheim, Norway. Email: astrid.lagreid@ntnu.no.

Martin Kuiper is Professor at the Department of Biology, NTNU. He develops tools and approaches for knowledge management and systems biology, with a focus on reaching out to biologists as users of this knowledge. Kuiper teaches systems biology introduction course for master students.

Address: Professor, Institute of Biology. Norwegian University of Science and Technology/ Norges Teknisk-Naturvitenskapelige Universitet (NTNU), 7049 Trondheim, Norway. Email: martin.kuiper@ntnu.no.