

UCLA

Publications

Title

Of domains, their knowledge, and their infrastructure

Permalink

<https://escholarship.org/uc/item/6sq6t87c>

Author

Ribes, David

Publication Date

2020-02-26

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Of domains, their knowledge, and their infrastructure

David Ribes

— *'Knowledge does not keep any better than fish'* Alfred North Whitehead

In my recent writings I have focused on the actors' category 'domain', often used in computational circles (eg, data science, AI, etc.), as with 'domain expert', 'domain scientist' or 'domain knowledge'. While domain is a term used casually – or vernacularly – I have argued that it is a keyword for computational fields: that it names a long-standing object of inquiry and of intervention, used today in a wide variety of interlocking fields such as librarianship, data management, knowledge management, and cyberinfrastructure, but that its genealogy can be, in part, traced back to early artificial intelligence. **The use of domain, far from innocent, activates a particular view of the role of computation and its relationship with 'the domains' (ie, everything else), and also plays that role in infrastructure circles (ie, infrastructure for this domain, that domain, or all domains).**

The concept of a domain is intricately entangled with the idea of knowledge. Domain is an old word, and we can track its etymology back centuries with a meaning that roughly coheres with its use today: a circumscribed sphere of human expertise or knowledge. But it is only beginning in the 1960s and 1970s that 'domain' gained its complement, that is: artificial intelligence, computer science, information science, or, today, data science. For instance, during that time, expert systems researchers called themselves 'knowledge engineers', and described themselves as 'capturing domain knowledge' in the formalisms that are the predecessors to today's computational ontologies and knowledge graphs. In relation to the domains, occasionally these fields describe themselves as 'domain independent', 'domain agnostic', or 'domain general', amongst other namings. Most often, no such naming is used at all, even while the relationship I am noting remains observable, as with the Venn diagram in fig.1. **These terms and images mark computational fields as something other than a domain, and more importantly, positions them as relevant to all the domains, ie, a universal or meta science.**

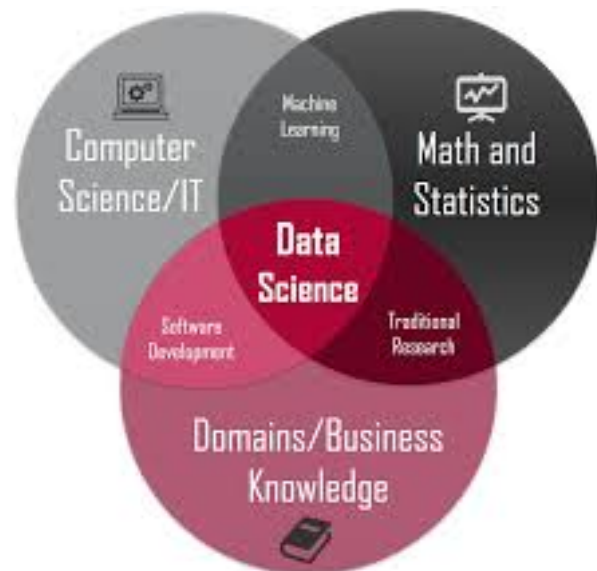


Fig.1 In perhaps the most common figuration for data science, this Venn diagram presents that field as the intersection of computing, statistics and domain knowledge. But note: one thing is not like the others. While computer science and statistics can be roughly mapped onto distinct fields, departments or professional bodies, domain has no particular referent, and may refer to essentially anything else, from geology to physics, but also, from travel agents to supply chains. Note also, it is the domains that have 'knowledge', and that without the domains and their knowledge, there would be no data science.

The concepts of domains and knowledge are intricately entangled with the idea of infrastructure, and this was the case long before I collaborated in writing the Knowledge Infrastructures report that animates this workshop. This is perhaps my strongest theoretical and methodological reservation to this coupling. For while in the report, coupling infrastructure and knowledge purported something new, in fact it harked to something quite old. More than just naming something old, I worry that it partakes in reproducing the very vision **I feel that we should inspect most carefully: the relationship between the domains and computational fields, or even the very parsing implied by casting these as two distinct kinds from their beginnings.**

We can observe the relationship I am trying to get at by turning to the cyberinfrastructure architecture diagram from the 2003 Atkins Report. In this stack diagram, “up” are the discrete and heterogenous domains, “down” is a singular cyberinfrastructure undergirding and connecting the domains. While figured differently than data science’s Venn diagram, it expresses the same idea: one set of fields (computation) will serve as the infrastructure for all the domains of science.

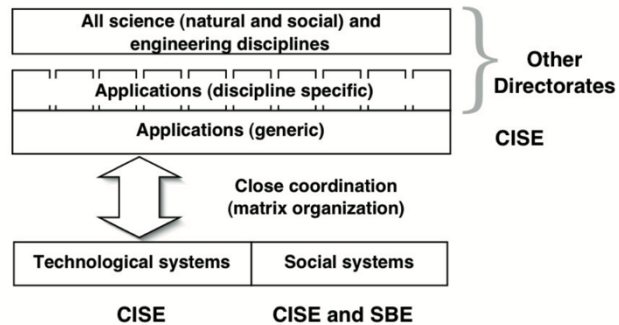


Fig.2 Cyberinfrastructure architecture stack, drawn from Atkins Report. "Up" are the heterogenous domains. "Down" is the unifying infrastructure. "All science" is the responsibility of the "other directorates", and the infrastructure is the responsibility of computer and social science (CISE and SBE).

The prompt for this essay asked us to discuss a dwindling infrastructure. The dwindling infrastructure of this essay is the vision for cyberinfrastructure itself, which while still active in science funding agencies, no longer operates as the marquee goal for NSF, NIH, or DOE, as it once did 2003-2013. Today data science is that marquee. My impressionist ethnographic take from the data scientists I work with today is that they find the vision of cyberinfrastructure, and its stack, to be “heavy”; overly reliant on standardization and centralization; too “top down”. Their technological ethos instead emphasizes agility, modularity, interoperability, decentralization, and ‘infrastructure as a service’ (ie, cloud). They tell me it’s a very different vision, and that, more than a vision, it speaks to a differently organized architecture.

And yet, something has persisted across these two imaginaries for scientific infrastructure: figure 1 and figure 2, both retain a vision that computation will act as the obligatory passage point for novel, data-intensive, and reproducible science. This has been my recent research interest, and I have found that **cyberinfrastructure and data science are only the tail-end of a long effort to place information, data, computation and their associated science&engineering fields, at the center of a well-funded sociotechnical enterprise for the rearrangement of the sciences (and beyond).** In this, we should take the coupling of knowledge and infrastructure, not as a natural or inevitable outcome, but as the object of inquiry. How did knowledge come to have an infrastructure?