# **MAGEE / ROSSITER — Statistical**

<0 figures>

# **Statistical Institutions and Database Disasters**

Magee, L. and Rossiter, N.

Arguably the relational database has had greater impact on the transformation of organizational cultures and the world economy than the Internet. The materiality of data centres coupled with the computational analytic potentials of databases has produced models of this world without historical precedent. Key here is the question of scale. The knowledge once derived from the transitional technologies of cabinets of curiosities (*Wunderkammer*), demographic registries, and Foucault’s ‘great tables’ in the 17th and 18th centuries—later systematised into various epistemic instruments that included Diderot’s encyclopaedia, the periodic table, the museum, and Linnaeus’ taxonomies—were all coincident with the rise of populations governed as statistical subjects. Such instruments can today be understood as proto-databases, foreshadowing what Gernot Böhme (2012) has called our present era of ‘invasive technification’.

The advent of the relational database in the early 1970s marks a particular turning point, we argue, in the ductility and malleability of knowledge concerning technical operations and the governance of society at large. At this point information becomes in a new sense purely programmable, and available for, among other things, forms of ad hoc knowledge production. The logical interrogation of subjects is literalised with the advent of ‘structured English query language (SEQUEL)’ in 1974, which presents as its very first example a resulting ‘relation describing employees’ featuring the barely fictional cast of ‘Jones’, ‘Smith’, and ‘Lee’ (Chamberlin and Boyce, 1974). Thereafter the relational model and its lingua franca, SQL, make possible for the first time entirely new fields of data science: data mining, informatics, business intelligence, real-time analytics. This in turn has led to a technological shift in the processing and logistical operations of modern institutions, with transformative effects in the apparently mundane fields of report writing, insurance assessment, credit checks, and policy development. What were once specialised arts become template-driven forms of analysis and institutional processes designed for replication. Here, knowledge rubs up against the politics of parameters. New uses of data become a constant in the social life of institutional settings, laden with a politics that remains for the most part implicit as it is pervasive. The durability of knowledge practices is coextensive with the persistence of parameters. Political existence contracts into the embodiment of Quine’s dictum: to be is to be the value of a variable.

By the early 1980s, the increasing reliance of all institutions on the relatively hypostatised *form* of the database reinforces and reinflects early-20th-century theories of institutionalism. For Weber, the institution was a necessarily constrained artefact of capitalist modernity. In the early eighties, DiMaggio and Powell (1983) revisited the ‘iron cage of bureaucracy’, reconceiving the modern institutional form as an ‘isomorphic’ entity with shared common procedures, structures, and operational norms while at the same time capable of adaptation to geographic, commercial, and industry-specific conditions. We argue that this isomorphism is recognisable by new institutionalist theorists in part due to its historical coincidence with the ubiquity and relatively enduring quality of the enterprise database, already emerging as a necessary part of modern institutional infrastructure. In the same way, the onset of flexible modes of capital accumulation was not a transformation independent of emergent developments in computational architectures. For example, the logistical world of ‘supply chain capitalism’ (Tsing, 2009) has become increasingly governed by the dual and interconnected processes of real-time computationally and just-in-time modes of production and distribution. The agility of the modern institution is, then, contingent upon the combinatory possibilities of relational databases that operate at ever increasing scales. The capacity for institutions to adapt to regimes of flexibilization is augmented, rather than replaced, by novel non-relational systems.

This twin model of data-and-institutional organization extends to the governance of logistical populations. One key example can be found in the recent challenges faced by the Australian Bureau of Statistics, which in 2014 confronted a very public institutional crisis of legitimacy based on a perception of computational failure. This crisis was precipitated by the multiplication of sites and points of data agglomeration: the ‘monopoly of knowledge’ enjoyed by the ABS for many years has now become rivalled by a diversity of institutional actors who also have considerable computational capacity to produce knowledge that bears upon how economies and populations are understood. The era of distributed computing, of virtualised clusters of machines and software that can co-operate to resolve queries over structured data on heterogeneous network and computational topologies, have been paralleled by questions of the sovereignty of singular guardians of population data. Since the 2000s, an array of new paradigms for arranging, connecting, and querying data—the Semantic Web, Linked Data, service-oriented architectures (SOA), and software-as-a-service (SaaS)—continue to bring into question claims over institutional legitimacy.

In the Australian context, the cutbacks in the operating budget of the ABS from successive governments needs to be seen in the context of the marketization of grey literature enabled by computational processes. This is not only a case of the state increasingly outsourcing a once-sacrosanct responsibility to private service providers. The multiple diffusions and aggregations of population data throughout a heterogeneous computational and institutional network means that the ‘database’ is no longer physically or conceptually containable within the borders of a single institution. The increasing dependency by policy makers on the generation of numbers by machines is symptomatic of the automation of decision making. Such is the institutional over-reliance on the pure power of computation. No matter how many manual double-checks and regulatory procedures may constitute the repertoire of techniques deployed to guard against the sort of institutional risk exposed by the ABS debacle, the scale and distribution of computational calculation in the production of knowledge will most likely result in an increasing jostling for legitimacy among institutional actors seeking government contracts related to policy development. Implicit in this jostling is a challenge to assumptions of ‘closed worlds’ and non-monotonicity that accompany the traditional relational database form and, by association, the single institution that manages such infrastructure. Rival claims, multiple perspectives, and contradictory or indeterminate datasets form the new territory of informational contestation. The case of the ABS offers an optic through which to view the emergence of such struggles.

Our claim is that this is less a story about decentralization and privatization of government within a neoliberal paradigm (although these are without doubt key forces), and more an instance of the technical logic of databases and distributed computing resulting in an unsettling of modern institutional authority. The relatively short history of the relational database has been mirrored, early on, by a recognition of the homologous or isomorphic character of the institutions that deploy them—and, more recently, by a sense that this isomorphism has itself ‘morphed’ or adapted into outright competition in the economy of data. Our central interest in this paper is to consider the role of the database as a technology of governance and the scramble of power as it relates to a capacity to model the world and exert influence upon it. What are the implications for public institutions as they relate to the supply of knowledge on national populations when the technologies of insight have become distributed and increasingly unaccountable across a range of actors? And what affordances does this present for the disruption of parametric politics, or the establishment, at the very least, of alternative parameters though which political life can be constituted?

**References**

**Böhme, G.** (2012). *Invasive Technification: Critical Essays in the Philosophy of Technology.* Shingleton, C. (trans.). Bloomsbury, London.

**Chamberlin, D. D. and Boyce, R. F.** (1974). SEQUEL: A Structured English Query Language. *Proceedings of the 1974 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control*. ACM, pp. 249–64.

**DiMaggio, P. J. and Powell, W. M.** (1983). The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields. *American Sociological Review,* **48**(2): 147–60.

**Tsing, A.** (2009). Supply Chains and the Human Condition. *Rethinking Marxism,* **21**(2): 148–76.