

Science, Technology, & Human Values

Semantic Web Practices: infrastructural politics and the future of the Web

Journal:	<i>Science, Technology, & Human Values</i>
Manuscript ID	ST&HV-2022-10-256.R3
Manuscript Type:	Original Article
Keyword:	semantic web, digital infrastructure, platforms, symbolic AI, interdisciplinarity, autoethnography
Abstract:	<p>In the past thirty years the Web has developed from its inception as a layer of protocols on top of the internet, to use by more than 5 billion people and organizations. This has driven the creation of vast quantities of data and led to deep concerns about the politics of digital data and computational methods. To date, critical investigation of these concerns has focused on large commercial platforms built on top of the Web, and their use of machine learning methods. Meanwhile, less attention has been paid to the underlying design and protocols of Web itself, and how these might be implicated in the very same process and concerns. We explore ongoing endeavors to transform the Web from a library of documents intended for humans, to a "semantic Web" using symbolic Artificial Intelligence to enable machine reasoning across multiple heterogenous data sources. In principle this would transform the production and circulation of knowledge at Web scale. We present the findings from an experimental, interdisciplinary study exploring the epistemological politics and sociomaterial practices involved in situated accomplishment of the semantic Web. Our findings have consequences for the future of the Web and the future of Web-based platforms.</p>

SCHOLARONE™
Manuscripts

Introduction

The past thirty years have seen the remarkable growth of the World Wide Web. From its inception as a new layer of protocols on top of the internet, the Web is now used by 5 billion peopleⁱ and is embedded in economic, social, and political life around the world. It is, by any account, an infrastructure for the 21st century, “as crucial to modern life as electricity, telephones and sewers” (Plantin et al. 2018, 10). In turn, the rapid growth of the Web has generated digital data at an extraordinary rate and scale, fueling a new round of Artificial Intelligence (AI) and raising deep concerns about the politics of digital data and computational methods. To date, the focus of these concerns has been the new “platforms” built on top of the Web, particularly large commercial organizations such as Facebook, Amazon and Google. Indeed, a sharp distinction is made with the non-proprietary, distributed and open qualities of the “original” Web, sometimes represented as a “public commons” for users. In contrast, the new platforms draw users into proprietary walled gardens where their activities leave digital traces, which can be commercialized through machine learning methods. Indeed, it is sometimes suggested that as these platforms continue to grow, they will rework the “open Web” putting its very future at risk (Helmond 2015; Plantin et al. 2018).

Our starting point is that this distinction between the open Web and platforms obscures critical questions about the Web as infrastructure and questions about the Web as infrastructure, and more specifically to the politics of data and computational methods. In this account, the Web sits in the background as a neutral enabler for platforms, which are the primary focus of attention. From an infrastructural perspective the fact that it is taken for granted is significant. As Jackson et al. (2007) note, the most effective infrastructures come to “appear as timeless, un-thought, even natural features of contemporary life” (Jackson et al. 2007, n.p.). However, this invisibility belies the activity that goes into producing such infrastructures, which are never fixed or finished “things” but made and re-made in ongoing sociomaterial practice. As such, infrastructures

1
2
3 embody values and politics, the outcome of choices shaped by debate and controversy (Bowker
4 et al. 2010), carrying these choices forward, often with profound consequences for what they do,
5
6 and do not, enable (Winner 1986; Star 1999).
7
8
9

10
11 This paper takes the Web as the focus of infrastructural inquiry. This has particular importance
12
13 at present, as long-standing ambitions gain ground in transforming the open Web from its
14
15 original form into a semantic Web. While the Web was originally conceived as a network of links
16
17 between documents and intended for human readers, a semantic Web adds structured machine-
18
19 readable meaning to the data inside documents. Enormous quantities of data are published on
20
21 the open Web, from a vast array of sources, on all subjects: far more data, and more diverse data
22
23 than in any of the new platforms. A semantically enabled Web would extract these data from
24
25 their documentary siloes and—taken to its logical conclusion—would enable machine reasoning
26
27 across all data published on the open Web. In contrast to the machine learning AI that has been
28
29 the focus of critical attention in studies of platforms, this application of symbolic AI would—in
30
31 principle—transform the Web into a single linked database for computational analysis.
32
33

34
35 Proponents' claims are underpinned by a positivist epistemology where data might be readily
36
37 modelled and linked for the greater good. However, this positivist epistemology belies the
38
39 infrastructural work required to achieve a semantic Web—particularly the epistemological and
40
41 ontological work required to create semantic linked data—and its consequences. To critically
42
43 examine the infrastructural work involved in building the semantic Web, this paper presents the
44
45 findings from an interdisciplinary empirical experiment designed to apply semantic Web
46
47 techniques to social research questions, specifically related to social class, aging and health.
48
49
50

51
52 How this work is done matters, with significant implications for the global production, control
53
54 and circulation of knowledge. What is at stake is how knowledge is represented and circulated in
55
56 the largest information construct in human history. However, shifting analysis from a high-level
57
58 recognition of what is at stake to empirical evidence of how these processes play out is a
59
60

1
2
3 challenge. This is technically complex work, not usually open or readily amenable to critical
4
5 sociological investigation. Directly addressing this challenge, we report on an interdisciplinary
6
7 and experimental research project, designed to explore the “situated accomplishment” of the
8
9 semantic Web as “a grandiose theoretical concept” (Woolgar and Lezaun 2013, 322) by tracing
10
11 “the fluid and unstable processes” (Van Heur et al. 2012, 342) through which semantic linked
12
13 data is produced. Overall, our paper un-freezes the sociomaterialities of semantic Web as an
14
15 emergent infrastructure. This has significant consequences for the future of the open Web,
16
17 because online data are described and modelled by tools which require conventions and formats
18
19 to be put in place, unlike the original permissionless Web (Halpin and Monnin 2016). It is also
20
21 significant for the future of the platforms that have dominated debates on data and AI to date
22
23 not least because the open Web provides the underlying infrastructure for these platforms
24
25 (Mukherjee 2019). Further, the major corporate platforms themselves are also investing in
26
27 semantic Web technologies (Halpin and Monnin 2016). Relatively little is known about these
28
29 highly sensitive commercial operations, but our paper provides some insights into the potential
30
31 consequences of semantic Web technologies for data and the knowledge that is derived from
32
33 them. We show the importance of understanding the sociomaterial practices involved in creating
34
35 the Semantic Web.
36
37
38
39
40
41

42 We begin by discussing the Web as infrastructure, the recent shift in attention from the original
43
44 Web to digital platforms, and how this shift raises critical questions about the ontological and
45
46 epistemological politics of digital data and artifacts. We then introduce the semantic Web and its
47
48 knowledge practices. Following an overview of our methodology, the paper focuses on the
49
50 empirical material, identifying five key junctures in the project, which we use to investigate wider
51
52 issues involved in creating infrastructure, including its consequences from both computational
53
54 and social research perspectives. We conclude by reflecting on the significance of these findings,
55
56 and widen out to consider how these same processes might shape the semantic Web at scale,
57
58 including implications for the future of platforms.
59
60

From the open Web to Platforms

The history of the Web is often framed in public discourse as an “invention,” originating in the proposalⁱⁱ for a new “global information sharing system” made by Tim Berners-Lee in 1989.

However, like all infrastructures, the Web did not appear *de novo* but was a response to the limitations of the internet and pieced together from existing technical artifacts. The internet had been steadily expanding the physical network of linkages between computers since the 1960s, but file transfer protocols remained complex, cumbersome, time-consuming and a specialist/expert preoccupation. By contrast, Berners-Lee’s proposal combined a simple, standardized document format, building on an existing markup language (SGML), with a hypertext and hypertext transfer protocols (HTTP) built on top of the internet. Using a “uniform resource locator” referencing the internet domain name system, documents could be published and accessed remotely regardless of bespoke operating systems and software.

The early 1990s saw the gradual enrolment of individuals and organizations running their own Web servers, using the new protocols and growing the nascent Web infrastructure, for example through the launch of effective search engines, standardization of protocols through the World Wide Web Consortium (W3C), and co-alignment with internet governance (Musiani 2015; Berners-Lee 2000; Brügger and Milligan 2019). Despite some early attempts to patent and commercialize the Web, over time it has been enacted and grown through an extensive network of actors, beyond the control of any individual corporation or government (Plantin and Punathambekar 2019).

In short, the Web is not a “thing” but a set of loosely orchestrated and dynamic sociomaterial practices. In 2005, the advent of social media sites led to the branding of Web 2.0 (Reilly 2005), initially distinguished from the original Web through the radical extension of user-generated content. Where Web 1.0 had enabled those with the skills and resources to publish a Website online, Web 2.0 enabled anyone with access to the Web to create online content through simple

1
2
3 interfaces. The growth in use that followed was extraordinary. Facebook acquired 315 million
4
5 accounts in its first five years (2004-9), reaching 1 billion by 2012 and just under 3 billion by
6
7 2023. Increasingly, these were not just Websites but a distinct form of commercial enterprise
8
9 built on top of the Web, now widely known as platforms (Gillespie 2010). Critical analysis of
10
11 these platforms is extensive, and centers largely on socio-political concerns about data and
12
13 computational methods. Detailed information on individuals and organizations becomes the
14
15 property of commercial companies, who aggregate and interrogate the data using machine
16
17 learning to construct new social forms of social analytics, marketing and targeting. This has
18
19 produced dramatic effects, for example the use of Facebook data by Cambridge Analytica as part
20
21 of the Brexit campaign in the 2016 referendum on membership of the European Union (Risso
22
23 2018). More insidiously, online behavioral tracking has become a standard commercial activity
24
25 (even with governance such as the General Data Protection Regulations in the European Union),
26
27 and sharing of individual data in the practice of everyday life is routinized.

28
29
30
31
32
33 At a more detailed level, critical platform studies pay attention to their sociomateriality . For
34
35 example, how the technical architecture of platforms shapes online socialities as well as the data
36
37 that are created and the forms of knowledge derived from these. The establishment of “friends”
38
39 and “followers” or “likes” for example, can have profound effects on social interactions and
40
41 identities (Van Dijk 2013) and has created artifacts for social media analytics, however
42
43 questionable their value may be (Halford et al. 2018). Following Bogost and Montford’s (2009)
44
45 injunction to take platforms seriously as computational infrastructures, Helmond (2015)
46
47 describes the computational work that drives platforms forward, showing how platforms are—
48
49 increasingly—extending their reach into the open Web, instating themselves on external
50
51 Websites (for example through widgets) and using a variety of methods to draw Web data into
52
53 their own data bases.
54
55
56
57
58
59
60

1
2
3 This is important analysis. Where we differ is in the characterization and positioning of the
4
5 “open Web” in the analysis of platforms. In these accounts, it appears almost as if the Web
6
7 “stopped” in the mid-2000s, and is now a relic from an earlier era, benignly enabling the rise and
8
9 growth of the new platforms, but now a passive prey to the inexorable rise of platform logic
10
11 (Holmes 2013), which taken to its logical conclusion will mark the demise of the Web (Plantin et
12
13 al. 2018). This stands in contrast to historical studies which emphasize the ongoing sociomaterial
14
15 production of the Web (c.f., Barnet 2019; Musiani and Schafer 2019). Extending this point, the
16
17 Web must be seen as a live, enacted infrastructure, rather than an ossified thing. Indeed, this
18
19 construction of the Web as static or superseded obscures its implication in some of the same
20
21 questions about the politics of data, AI and infrastructure as those raised in studies of platforms.
22
23
24
25

26 The Rise of the Semantic Web

27
28
29 In this paper we focus on a different and longer-term challenge to the original Web, specifically
30
31 ongoing challenges to its grounding as a document-based system, rather than a data-based
32
33 system. While vast volumes of data were published inside documents, the established protocols
34
35 offered no way of link across these data. A “semantic Web” would add machine-readable
36
37 meaning to data published online and use computational tools to support machine reasoning
38
39 across the Web. Machines would be able to take data from one source to complement or extend
40
41 data from other sources, combining information at Web scale “to make inferences, choose
42
43 courses of action and answer questions.”ⁱⁱⁱ To achieve this, a standardized knowledge
44
45 representation system was required to model data entities (people, places, things), their
46
47 properties, and their relationships. While the original Web, demanded only limited
48
49 standardization to make simple links between documents, the semantic Web demanded a “full
50
51 blown language for knowledge representation” (Halpin and Monnin 2016, 6) to enable
52
53 computational reasoning across Web data.
54
55
56
57
58
59
60

1
2
3 Like any infrastructure, the semantic Web has been subject to controversy and dispute. In the
4 semantic Web community, this has been characterized by two different approaches reflecting
5 longer term divisions on the philosophy of AI. The initial approach to building a semantic Web
6 was shaped by the declarativist tradition “built on a foundation of logical axioms that precisely
7 described the permitted inferences of any statement made,” and “first order logic that could
8 express any and all knowledge to be published on the Web” (Halpin and Monin 2016, 129).
9
10 From this perspective, a semantically enabled Web would depend on the development of
11 consistent and durable global semantics. This approach raises two key challenges. First, the scale
12 of knowledge engineering work required given that adding structured and consistent meaning to
13 data is a major undertaking. Doing so in a shared and consistent way demands co-ordination and
14 collaboration across multiple sites of data publishing, in order to realize interoperability. Second,
15 concerns were raised about the rationalization of a permission-less Web (Halpin and Monnin
16 2016) into a structured information system. What forms of knowledge could and would be
17 expressed by semantic Web technologies and what may that mean for the future of the Web?
18 (Shirkey 2006; Halford et al. 2013; Ford and Graham 2015; McCarthy 2017). The Web
19 infrastructure would no longer be indifferent to content but, rather, would depend on
20 prescriptive ways of describing entities and their relationships. These would be reproduced at
21 scale and—at the same time—obscure to those not familiar with their philosophical and
22 computational underpinnings. In this context, there were concerns about what would happen to
23 partial, inconclusive or context-sensitive information (Ford and Graham 2016) in the push for
24 global semantics. In contrast to a declarativist approach, a growing “linked data” community was
25 making use of semantic Web technologies for individual, localized projects marking a shift
26 toward a “proceduralist” philosophy. Rather than depending on generalizable statements,
27 proceduralists understand intelligence as know-how, where the tools for reasoning are
28 inseparable from domain knowledge, focusing on data infrastructures rather than logical
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 modelling (Winograd 1975). This is a more specific and partial approach driven by local
4
5 priorities, that values pragmatism and simplicity over formalism and consistency (Poirier 2017).
6
7

8
9 These two approaches to the semantic Web resonate with wider and longstanding distinctions in
10
11 Computer Science between “neat” and “scruffy” (Poirier 2018). This brings philosophical
12
13 distinctions down to earth, focusing on differences between theoretical and applied approaches
14
15 in computational practice. In short, neat is principled, consistent and interoperable; scruffy gets
16
17 something done, albeit in a particular context. Poirier (2018, 359) describes the history of the
18
19 semantic Web in the 2010s as a “turn for the scruffy,” linking this both to a proceduralist
20
21 position on knowledge representation, and in pragmatic recognition of the political-economic
22
23 considerations that shaped semantic Web activity, such as end goals, deliverables, funding, and
24
25 market demand. In short, existing accounts suggest that the semantic Web is emerging in
26
27 distinct, diverse forms. What we lack is empirical evidence of the consequences of these different
28
29 approaches for the forms of knowledge produced. In the remainder of this paper, we present
30
31 findings from an interdisciplinary experimental project designed to explore how the semantic
32
33 Web is “done.”
34
35
36
37
38

39 Methodology

40
41
42 Our project *Social Sciences, Social Data and the Semantic Web* was designed specifically to explore the
43
44 epistemological and ontological stakes of the semantic Web. In designing our research, we faced
45
46 the usual challenges of studying infrastructure, which by definition “typically sits in the
47
48 background...is invisible, and...frequently taken for granted....In such a marginalised state its
49
50 consequences become difficult to trace and politics are easily buried in technical encodings”
51
52 (Bowker et al. 2010, 98). Our methodology was designed to unfreeze (Star 1999) the semantic
53
54 Web by surfacing the invisible work involved in re/producing semantic linked data. Specifically,
55
56 our project was interdisciplinary, experimental and autoethnographic.
57
58
59
60

1
2
3 The project was designed by a team of two Computer Scientists and two Sociologists and
4 employed two researchers with recent interdisciplinary PhDs in social and computational
5 sciences. It had small funding^{iv} from a UK Economic and Research Council program [expressly
6 intended to support blue skies research. This gave us “the courage to try something new [in] a
7 permissive environment” that supported interdisciplinary research (Bijsterveld and Swinnen
8 2023, 2). The key research questions were framed by sociology, but it did not provide the
9 disciplinary expertise to untangle the technical complexities of semantic Web technologies. The
10 computer scientists were experienced with semantic Web technologies, but their training had not
11 prepared them for the epistemological or substantive questions raised by this project. For all of
12 us, this was without a doubt the most challenging project we had ever worked on, and the
13 riskiest by far as it fell way outside the usual norms and rewards of our home disciplines.

14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29 Working together we sought to explore how semantic Web technologies “materially organize and
30 instantiate relations between people, things, perspectives, and technologies” (Gray et al. 2018, 1).

31
32
33 All members of the team had worked together previously and shared a commitment to
34 interdisciplinary collaboration for critical examination of Web technologies. Our collaboration
35 had developed over time, through shared teaching and research activities that had led to “shifts
36 in how we read, value concepts, critically combine methods, cope with knowledge hierarchies
37 and adopt writing styles” (Bijsterveld and Swinnen 2023, 3-4). In Schubert and Kolb’s (2019)
38 terms, this was a symmetrical engagement in which Computer Science and Sociology brought
39 distinct and in-depth expertise to the project and neither was in the service of the other.

40
41
42
43
44
45
46
47
48
49
50 Our approach was collaborative experimentation (Balmer et al. 2015) based on converting two
51 conventional social data sets into semantic linked data, to make the implicit infrastructuring
52 processes visible. Including social scientists as participants in the process, rather than as external
53 observers of a Computer Science project allowed us to open-up challenges and resolutions that
54 would otherwise remain unarticulated (Balmer et al. 2015). Our approach was experimental in
55
56
57
58
59
60

1
2
3 two senses. First, we were agnostic regarding different approaches within the semantic Web. We
4 did not have a pre-defined approach for how the semantic Web should be done, our aim was
5 simply to trace how it happened. Second, the process itself was the object of inquiry rather than
6 the means to an end. The experiment was to see how this process evolved and it is this that we
7 concentrate on here.
8
9
10
11
12
13
14

15 The experiment was grounded in a specific case: creating semantic linked data for research on
16 social class, aging, and health. We chose this topic for two reasons. First, it is characterized by
17 conceptual disputes, complex research questions and highly dynamic data. This was a deliberately
18 tough test case for semantic Web technologies, intended to render taken-for-granted
19 infrastructural practices more visible than they may otherwise have been. Second, there were two
20 publicly available data sets available for experimentation: the Great British Class Survey (GBCS
21 2013) and the English Longitudinal Survey of Ageing (ELSA Wave 6 2012). The GBCS was a
22 pioneering study driven by Bourdieusian approaches to social class, which moved beyond
23 occupational categorizations to include economic, social, and cultural capitals as a basis for
24 defining a new class structure for the UK (Savage 2018). However, it had little information on
25 aging and health. The English Longitudinal Survey of Ageing was rich with information on age
26 and health, and included economic, social and cultural data, but relied on occupational
27 definitions of social class (NS-SEC). These are two separate surveys, and we cannot know if any
28 of the same individuals appear in both data sets. We did not set out to link information on
29 individuals. Rather, our aim was to critically examine the ontological and epistemological
30 consequences of using semantic Web technologies to infer GBCS social classes to explore health
31 inequalities in the ELSA population.^v
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53

54 Our approach was autoethnographic. We documented and reflected on our activities throughout
55 the project.^{vi} Data collection took place from January 2019 to January 2020. Three main types of
56 data were collected: (1) audio recordings and photographs from team meetings and the project
57
58
59
60

1
2
3 Advisory Board; (2) fieldnotes and photographs from technical meetings; (3) a full record of all
4
5 technical development in GitHub. Initial thematic analysis of the data was undertaken by the
6
7 lead ethnographer and discussed in depth at full-team data workshops. In this process, we
8
9 identified five critical junctures across the lifetime of the project where important
10
11 epistemological and ontological questions about our use of semantic Web technologies surfaced.
12
13 In what follows, we describe each of these junctures and open-up what was at stake for our
14
15 project. Because these insights are presented for a Social Science audience here, the conventions
16
17 of writing lean more in this direction. However, because our argument focuses on the materiality
18
19 of the semantic Web, we focus on technical detail to a degree that means what follows may be
20
21 less familiar for a Social Science audience. However, it is precisely this attention to detail that
22
23 allows us to unpack the situated accomplishment of the semantic Web, and to explore what is at
24
25 stake in these knowledge practices.
26
27
28
29
30

31 Creating Semantic Linked Data

32
33

34 In this section we explore five junctures during the project. All entailed debate about the best
35
36 way forward and involved both the Sociologists and the Computer Scientists. Although the focus
37
38 was always substantive—how best to represent the social research data—the choices centered on
39
40 differences between declarativist and proceduralist approaches, shaped within a wider landscape
41
42 characterized by local pressures and contingencies. In what follows we describe these choices
43
44 chronologically.
45
46
47

48 We began by thinking about how to represent social class and aging in a machine-readable way.

49 In the semantic Web, “ontologies”^{vii} provide a formal description of the classes of object, their
50
51 properties, and relationships in a field. The *first key juncture* concerned whether to create a
52
53 conceptual ontology of social class, aging and health, or an applied ontology derived from the
54
55 data sets we planned to use. The former aligns with a declarativist perspective, using logic to
56
57
58
59
60

1
2
3 model definitive statements about what can exist. The latter aligns with a proceduralist approach,
4
5 using a specific artifact (survey instruments in this case) to create a model of the field.
6
7
8
9

10 In Computer Science:

11
12
13
14 What we mean when we talk about an ontology is an engineered artifact, it is a shared
15
16 conceptualization of some domain or area of study or area of interest. (Computer
17
18 Scientist 2, Advisory Board Meeting notes, 15 March 2019)
19
20
21

22
23 This can often involve importing established schema from outside the semantic Web:
24
25

26
27 On other projects I've worked on it was "here's your list of seed types." It's not
28
29 something you can argue about...it just exists, it's schemas with pre-existing hierarchies
30
31 and you just have to convert it into semantic form, whereas we spent a lot of time
32
33 discussing...how best to represent the objects in the ontology. (Computer Scientist 3,
34
35 Team Meeting, 22 March 2019)
36
37
38

39 And there were already identified data sets for the project:

40
41
42 As soon as I saw the data sets and realized that they needed to be linked together in
43
44 some way, my assumption went down the data route... maybe that's different from the
45
46 original view of the semantic Web (Computer Scientist 3, Team Meeting, 22 March 2019)
47
48
49

50 From a sociological perspective, the means to construct an ontology were less clear:

51
52
53 We could have analyzed the data in different sorts of ways, there was a fight in the
54
55 team... [and] of course there are other people who may have done it differently.
56
57 (Advisory Group Member 1, Advisory Board Meeting, 15 March 2019)
58
59
60

1
2
3 Collectively we were all acutely aware of outputs committed to the funder, and the tight
4
5 deadlines. An applied ontology was easier and quicker than a conceptual ontology. This was a
6
7 scruffy decision: driven by pragmatic needs rather than taking a principled approach to create a
8
9 comprehensive ontology which would have been “neater” and (likely to be) of greater use
10
11 beyond our project. The critical consequence of this decision was to narrow the representation
12
13 of the domain to a particular artifactual source, deferring key decisions about what was in the
14
15 ontology to the GBCS survey (as we modelled this first). This meant the original theoretical and
16
17 methodological approaches in GBCS became hardwired as “the field,” rather than part of a
18
19 contested field. We also inherited some questionable classifications from GBCS (for example, a
20
21 binary representation of gender); and a specific set of activities that were important from a
22
23 Bourdieusian perspective (for example, social engagement and cultural consumption). Once
24
25 these decisions are represented in the ontology, they implicitly restrict what is knowable
26
27 whenever the ontology is used, whether for its original purpose or as it is repurposed by others.
28
29
30
31
32
33 Once we made the decision to base our ontology on GBCS, *a second key juncture* concerned debate
34
35 over *how* to model the ontology. Initially, we began by organizing this around cultural, social and
36
37 economic capitals, based on the structure of the GBCS survey, intending to pick out the implicit
38
39 semantics for the ontology (informed by reading the underpinning sociological research). This
40
41 proved difficult. SW technologies lacked the expressivity to capture theoretical nuances:
42
43
44

45 I think those things [social class] are contested among sociologists and the danger is that
46
47 if we model that as a top-level thing, somebody else is going to come along and say “well
48
49 actually I think *that* should be cultural capital and *that* shouldn’t be”...I’m worried that
50
51 some of the terms we are using here...are a bit more contested. (Computer Scientist 1.
52
53

54 Team Meeting notes, 3 April 2019)
55
56

57 By sticking to the GBCS questionnaire, we risked reifying decisions already made by social
58
59 researchers in the 2010s. Not only definitions of social, economic, and cultural capitals but also
60

1
2
3 tying survey responses (facts, preferences, activities) to these categories. In turn, this raised issues
4
5 of interoperability:
6
7

8 In looking for GBCS's measures of cultural capital in ELSA, I have encountered various
9
10 dilemmas...about if and how variables could be considered relevant to GBCS measures
11
12 of cultural capital. (Social Scientist 2, email, 30 April 2019)
13
14

15
16 If the ontology was to be useful for different users and contexts, a more generalized way of
17
18 representing knowledge that didn't concretize theoretical positions derived from the GBCS
19
20 survey was needed. We agreed on a radical simplification, removing social, economic and cultural
21
22 capitals as the organizing features of the ontology, separating out the underlying theory and the
23
24 survey instrument as design principles. We chose to use the more generic classes of "person"
25
26 and "activities" as the common denominator. Effectively, the ontology became distanced from
27
28 explicitly Bourdieusian class analysis to provide the (apparently) least contested description of
29
30 the data. This made it easier to link across our two data sets—since we knew that ELSA did not
31
32 have the same terms and measures as GBCS. It also meant that this ontology could be used
33
34 more widely by others.
35
36
37
38

39 So, you take a data set, in this case GBCS, and represent it in such a way that we could
40
41 use that representation for the job of data integration between different data sets so we
42
43 could do reasoning with it. (Computer Scientist 2, Team Meeting notes, 21 May 2019)
44
45

46
47 For better or worse, this represented a shift back from the procedural approach—modelling the
48
49 data—to a more declarativist position, making definitive claims about what entities exist (person,
50
51 activity) rather than modelling a survey instrument. A more declarative approach opens the
52
53 potential for re-use. The ontology no longer had any explicit Bourdieusian theoretical framing,
54
55 nor contextual detail about its origins. It now defined facts about people (income and other
56
57 sources of wealth, for example), as well as things that people do (activities) but not from an
58
59 explicit theoretical position. By now, this is not an ontology of class but an ontology of facts
60

1
2
3 about individuals and the activities they engage in—albeit still based on GBCS—that may be
4 used for a variety of purposes. By not imposing rigid categorical or conceptual decisions we leave
5 it open to others to overlay their views or categorizations on top of the base ontology. A
6
7 consequence of this is that information about the Bourdieusian framing of the data collection
8
9 becomes lost as categorical aspects of the instruments are removed from the modelling.
10
11 However, re-use may require the extension of the base ontology, should there be requirements
12
13 for a broader set of activities to support different ways of interrogating the data.
14
15
16
17
18

19 Once we had a basic ontology design, drafted the *third key juncture* was about how to model time.
20 This was in part prompted by the longitudinal nature of the ELSA data set, but more generally
21 by the semantic Web imperative to create resources that can be re-used and re-purposed over
22 time. *When* a fact was recorded (for example, house price or income) is important. With the
23 declarative emphasis on standardized and consistent models, we sought to use an existing
24 ontology. However, those created for describing temporal concepts (topological ordering,
25 instants or intervals) were too specific, and OWL-Time—the declarative standard for the
26 semantic Web—was more extensive, demanding a time stamp for each piece of data: “I mean
27 OWL-Time is a bit [heavy]... it is not quite clear what all the things in OWL-Time do.” (Computer
28 Scientist 2, Team Meeting, 21 May 2019).
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 This made it very costly computationally, at the expense of scale and speed. Instead, we turned
46 to the smaller yet declarative “fluents” ontology, created to manage inference across key aspects
47 of time in a computationally efficient way (Welty and Fikes 2006). This ontology is declarative,
48 neat and formal (based on situational calculus). Adopting it meant reconfiguring the ontology,
49 adding temporal parts for every individual and allowing a common timestamp for data collected
50 in a single survey, rather than an individual timestamp for each piece of data. A range of other
51 temporalities may also appear, including frequency (of particular activities for example) or
52
53
54
55
56
57
58
59
60

1
2
3 duration (periods of employment or residency, for example). Adopting this modelling approach
4 established critical path dependencies for later in the design process, as explained below.
5

6
7 Decisions on modelling time had consequences for the tractability of types of temporal queries
8 or how that data might be consumed by larger systems with different overarching temporal
9 models.
10
11

12
13
14
15 The *fourth key design juncture* occurred at the point of re-introducing social class into our toolkit. By
16 taking Bourdieusian capitals out of the data-based ontologies we had simplified and generalized
17 them to support re-use by other researchers who could now query the data in a highly flexible
18 way. However, we *were* interested in Bourdieusian capitals and the GBCS social classes (and we
19 suspected that many other social researchers would be too). Specifically, we wanted to create the
20 computational means to derive GBCS social classes from ELSA (as a test case), which would
21 then allow us to integrate GBCS classes with analysis of health and aging. We were also
22 interested in overlaying different understandings of class on top of existing data sets, in our case
23 comparing GBCS with NS-SEC^{viii} to examine health inequalities among older adults. Given our
24 previous decision to model the data and not the domain, and then to focus on “activities” as the
25 core of our ontology, we now had to create a *separate* ontology for social class. Both the GBCS
26 and NS-SEC class models proved easy to model declaratively, drawing on the underpinning
27 research to define the different class hierarchies conceptually. But to produce actual class GBCS
28 dispositions from ELSA we needed to engage with empirical data.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46
47 The semantic Web tool of choice for this is inference, which gives machines a set of abstract
48 rules to work out knowledge they don't have from knowledge that they do have. That meant
49 providing rules to derive GBCS social classes from capital measures (social, cultural and
50 economic) in a data set (ELSA) that did not use this model of class analysis. These rules are
51 written inside the ontology and published along with it, to make them discoverable and usable by
52
53
54
55
56
57
58
59
60

1
2
3 other machines to interrogate and examine other data sets. We modelled the inference rules from
4
5 the sociological research (Figure 1).
6

7
8 [INSERT Figure 1 here]
9

10
11 So far, so neat. However, while inference works with logic, some forms of inference need
12
13 statistical information about the data set (e.g. categorical thresholds). The symbolic approach to
14
15 AI embedded in semantic Web technologies is not well suited to complex mathematical
16
17 calculations, such as inferring class membership based on values measured against thresholds.
18
19 These processes involved knowledge that did not reside within the data sets:
20
21

22
23 In the descriptions for economic capital for seven classes, words like “very high” and
24
25 “high” are used, but do you have any qualifiers for this? Currently my approach was to
26
27 assume the numbers in Table 6 were an average of the scores and then fit the categories
28
29 around that, but obviously that will only work if those numbers are the average. (CS1,
30
31 questions to GBCS team, 7 June 2019)
32
33

34
35 Our pragmatic solution was to supplement our inference rules in the ontology with an algorithm
36
37 informed by empirical evidence from the findings of the GBCS to calculate the capital measures
38
39 for individuals that the inference rules would then use to reach a class disposition. This
40
41 supplements the declarative and neat inference rules that would allow machines to make
42
43 deductions, with proceduralist and scruffy algorithms providing a specific set of instructions that
44
45 tell machines exactly what to do with a specific data set when directed to do so. The thresholds
46
47 used in these algorithms were derived for the specific dataset through analysis, and this
48
49 knowledge resides external to the semantic Web data and would require generating for any new
50
51 data set, placing further limitations on re-use.
52
53

54
55
56 So far, we have described the preparatory work involved in creating a package of tools to
57
58 support sociological analysis of GBCS social classes. We have created three ontologies, some
59
60

1
2
3 inference rules, and an algorithm. Following the descriptions in the two ontologies, we used
4
5 R2RML^{ix} to convert the GBCS and ELSA data sets from their spreadsheet forms to semantic
6
7 linked data (transforming the terms and structures in the datasets into their ontology
8
9 equivalents). The final challenge was to run queries across the data sets. Our proof of concept
10
11 was to derive GBCS class dispositions for ELSA using inference rules. We tested this using a
12
13 small sample from GBCS, to see how our methods performed compared to the original class
14
15 recorded there, and it performed well when comparing our ascribed class to the most likely class
16
17 generated by the GBCS K-means clustering (Savage et al. 2013). However, when we ran queries
18
19 across larger volumes of data in ELSA, the process became intractable. Our interpretation was
20
21 that the fluents approach to time caused this problem:
22
23
24

25
26 The Fluents ontology has had a great impact on the project, I'm still convinced that's
27
28 why the reasoner doesn't work...It created a lot of extra triples^x and I'm still convinced
29
30 that's the reason that the Reasoner kept breaking. (Computer Scientist 3, Team meeting
31
32 transcript, 15 October 2020)
33
34
35
36
37
38

39 We've got a very semantic Web approach which means we could do lots of reasoning, if
40
41 only the reasoning scaled! Could we have come up with a simpler approach? Yes, we
42
43 probably could. Would it allow us to do the same things [given compute capacity]? It
44
45 wouldn't. (Computer Scientist 2, Team meeting transcript, 15 October 2020)
46
47
48

49 Although it had been designed to be an efficient declarative statement of time, there is no record
50
51 of the Fluents ontology being used to query across even moderate-sized data sets. While this had
52
53 seemed to be a declarative and neat solution earlier on, now it was challenging the whole
54
55 outcome of the experiment. The semantic Web is better suited for some things than others.
56
57 Complex numerical calculations or modelling temporal data are challenging. As a result, they are
58
59 often avoided, or dealt with outside of the semantic Web, reducing data interoperability.
60

1
2
3 Consequently, the *fifth juncture* concerned whether to persevere with inference^{xi}—demonstrating
4 the declarative power of the semantic Web—or move to using algorithms outside of the
5 ontology, to provide a specific solution for our project. Given the pressures of time, staffing, and
6 limits of the reasoning technologies, we chose to devise an additional algorithm—deriving class
7 dispositions from capital measures—to achieve our proof of concept. This was more
8 proceduralist, and scruffy approach. It did not demonstrate the power of machine reasoning.
9
10 Although our algorithms are re-usable, they are not discoverable in the semantic Web because
11 they require specific execution to generate values and are not encoded in semantic Web schema.
12 They would also have to be re-written for different data sets and contexts. From a social research
13 perspective, this meant we now had the means to interrogate ELSA with knowledge from
14 GBCS. However, our (unusual) social scientific use of semantic Web technologies is
15 compromised to the extent that it is marginal to the (growing) semantic Web, not completely
16 discoverable by others or not part of the knowledge base. This means social researchers are less
17 able to benefit from the full computational opportunities of the semantic Web.

34 35 Discussion and Conclusions

36
37
38 The Web is now a core infrastructure for the modern world. Like any infrastructure, how it is
39 done matters. This paper takes the Web as the focus of critical infrastructural inquiry to explore
40 the politics of data and AI methods in the evolution of the open Web. These questions here are
41 distinct from those about the politics of data and AI in the big corporate platforms such as
42 Facebook, Amazon and Twitter. These platforms are centralized and proprietary, and critical
43 concerns focusing on the use of machine learning methods to extract information and meaning
44 from users' content and activities within the platforms. The emergent semantic Web uses
45 symbolic AI to generate a new layer of standardized tools and protocols for data linkage and
46 inference across multiple, diverse, distributed and open Web data sources, with a focus on
47 scientific and practical knowledge (Berners-Lee et al. 2001) rather than deriving information on
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 users' activities online. Nonetheless, the potential implications are profound. No less than a new
4
5 approach to how knowledge is represented, created and circulated across the open Web.
6
7

8
9 Our project interrogated the situated accomplishments of the semantic Web (Woolgar and
10
11 Lezaun 2013, 322) of the semantic Web, in one small instance, exploring the sociomaterial
12
13 practices of the SW and their consequences for the knowledge created. This is the first project to
14
15 explore this empirical detail, as far as we are aware. Overall, our project demonstrated two
16
17 central points. First, that while high-level debates about declarative and proceduralist approaches
18
19 have emphasized two very different approaches to the semantic Web (Halpin and Monin 2016;
20
21 Poirier 2017), such binary framings were neither accurate nor helpful in explaining how the
22
23 semantic Web was done in practice. In our experience, creating semantic linked data was not a
24
25 choice between declarativist and proceduralist approaches, or between neat and scruffy. Rather,
26
27 our experiment shows that these different approaches are contextual and relational. There was a
28
29 continual pull between the local demands of the project—intellectually and practically—and the
30
31 wider goal of producing interoperable and reusable artifacts and tools. In the end, specific
32
33 choices were made to embed local and specific ontological and epistemological decisions in what
34
35 would appear as a generic set of tools. Second, that the epistemic constraints of the Web demand
36
37 workarounds for certain kinds of data and querying. Symbolic AI is not well-equipped to deal
38
39 with quantitative data, so this work was outsourced to algorithms. Also the computational costs
40
41 of modelling time made these queries intractable. The significance of these points is that certain
42
43 kinds of data or queries may be marginalized by the semantic Web or hidden outside the
44
45 transparent and open structures of the semantic Web. More widely, the data and tools that we
46
47 created carry these specific ontological and epistemological choices and consequences forward, if
48
49 and when they are re-used and become part of the wider semantic Web infrastructure.
50
51
52
53
54
55

56
57 The promotion and publication of semantically linked data on the open Web is growing fast
58
59 (www.lod-cloud.net, <https://opendatacharter.net>, www.wikidata.org). By publishing data and
60

1
2
3 underlying ontologies as linked and open, the explicit intention is that others will re-use and re-
4 purpose them. Most effectively (for SW proponents) particular ontologies will become the single
5
6 standard for describing particular domains, both empirical fields and structuring processes such
7
8 as time or space. Our small study of the knowledge practices involved in creating semantically
9
10 linked data, and their consequences, highlights the importance of raising how these same
11
12 questions and processes might shape the semantic Web at scale. The data sets and ontologies
13
14 published on the open Web have also been produced through concrete knowledge practices,
15
16 shaped by the priorities and contingencies of the organizations and individuals involved. This
17
18 means they inherit and bear particular decisions about how to represent contested entities,
19
20 relations and fields. Linked data artifacts carry decisions about the meanings that are “wired
21
22 into” data sets and will—in principle—extend beyond their origins into any other context of use.
23
24 These decisions can be very difficult to excavate once they have been materialized into a
25
26 particular ontology (as we saw with the disappearance of any explicit reference to Bourdieu in
27
28 our ontology, albeit that the legacy remained implicit and unarticulated in some categories that
29
30 remained in the ontology). Extending the implications of our findings to other contexts where
31
32 semantic Web technologies are used, similar knowledge processes will shape those artifacts and
33
34 shape activities in any other contexts where they are re-purposed. This could be explored further
35
36 by interrogating the RDF data and underlying ontologies that are published in the linked data
37
38 cloud as part of the wider commitment to data sharing and interoperability.
39
40
41
42
43
44

45
46 Finally, the emergence of a semantic Web also has consequence for the big commercial
47
48 platforms and our understanding of these. There are two different points to be made here. First,
49
50 semantic linked data provides a new resource for emerging “infrastructuralized platforms”
51
52 (Plantin et al. 2018), that has been created on and by the open Web. The linked data cloud offers
53
54 open data that can be ingested by the platforms which will, in turn, inherit the epistemological
55
56 and ontological consequences of symbolic AI and the particular assemblages produced. This will
57
58 extend and reproduce the knowledge practices of particular SW projects/activities inside the
59
60

1
2
3 platforms, even if these platforms present as walled gardens (Holmes 2013), for example in the
4
5 hypothetical case that a major social media platform used Wikidata to construct biographies, it
6
7 would inherit the inherent classifications of the original ontologies. Researchers have pointed out
8
9 how platform engineering reaches out into the Web, for example through the use of widgets
10
11 (Helmond 2015), which confirms that SW engineering may shape platform activities. This raises
12
13 further interesting and important questions about the co-constitution of platforms and the Web
14
15 (Mukherjee 2019).
16
17

18
19 Second, there is evidence that platforms themselves are using SW methods for their own
20
21 operations, alongside more fully established machine learning methods (Poirier 2017). However,
22
23 we know little about how these methods are deployed inside platforms, for example which data
24
25 are used or how, what conceptualizations or models are in play. What we can reasonably suggest,
26
27 based on the findings or our own experiment and the wider platform studies literature, is that
28
29 these investments in SW technologies will be shaped by questions of local contingency and
30
31 expected to produce actionable knowledge that can be fed into platforms' corporate
32
33 assemblages, shaping how search engines work, how recommendations are made (e.g., on
34
35 Facebook feeds). In turn, this may instate epistemological and ontological effects of SW methods
36
37 into wider questions about fake news, filter bubbles and discriminatory decision-making tools.
38
39 Further, the epistemic limits of the semantic Web technologies we have explored in this study
40
41 (and likely other limits not yet apparent) will shape their use inside large corporate platforms, and
42
43 subsequent consequences. For example, the intractability of temporal reasoning described above
44
45 may narrow the forms of analysis inside platforms, and consequently what kinds of information
46
47 appears through these platforms, and therefore what can and can't be known through them. For
48
49 example, if Google is used to search for historical information and it draws on semantic Web
50
51 technologies to generate this data, there would be no way of knowing how these data were
52
53 constructed or modelled, nor the epistemic limits that may be embedded in the methods used
54
55 (ontologies or forms of inference, for instance).
56
57
58
59
60

1
2
3 Our key point is to focus on the sociomaterialities of the open Web as a dynamic infrastructure,
4 and pay attention to how these sociomaterialities are embedded in the ongoing politics of data
5 and AI. Attention to the emerging semantic Web raises critical questions about how knowledge
6 is produced online, by whom and with what consequences that both sit alongside and contribute
7 to more widely debated concerns about platforms, data and AI. Unpicking the sociomaterialities
8 of these knowledge practices is painstaking and challenging work that demands the integration of
9 expertise in social and computational sciences. As we have shown, the devil is in the detail—
10 understanding knowledge practices demands sociological engagement with technical processes
11 and computational openness. To understand knowledge practices is to question taken-for-
12 granted ways of doing. Even then, we can never expect to unpick every last detail. The key is to
13 generate sufficient knowledge of the kinds of processes and questions that should be asked to
14 ensure that the infrastructures of the open Web are not taken for granted or sink into the
15 background, and instead remain the object of critical investigation.
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

36 REFERENCES

- 37
38
39 Balmer, A. S., Calvert, J., Marris, C., Molyneux-Hodgson, S., Frow, E., Kearnes, M., Bulpin, K.,
40 Schyfter, P., MacKenzie, A. and Martin, P. (2015) 'Taking Roles in Interdisciplinary
41 Collaborations: Reflections on Working in Post-ELSI Spaces in the UK Synthetic Biology
42 Community', *Science & Technology Studies*, 28(3), pp. 3–25.
43
44
45
46
47
48
49 Barnet, B. (2019) 'Hypertext before the Web—or, What the Web Could Have Been' in Brügger,
50 N. and Milligan, I. (Eds) *The SAGE Handbook of Web History* London, Sage.
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

1
2
3 Berners-Lee, Tim, James Hendler, and Ora Lassila. "The Web Semantic." *Scientific American*
4 284, no. 5 (2001): 34-43.
5
6

7
8 Berners-Lee, T., Hall, W., Hendler, J., O'Hara, K., Shadbolt, N., Weitzner, D., (2006a) 'A
9 Framework for Web Science.' *Foundations and Trends in Web Science* 1(1) pp. 1-130.
10
11

12
13 Bijsterveld, K., Swinnen, A. (2023) *Interdisciplinarity in the Scholarly Life Cycle*. Palgrave
14
15 Macmillan Cham
16

17
18 Bogost, I., & Montfort, N. (2009). 'Platform Studies: Frequently Questioned Answers.' UC
19 Irvine: Digital Arts and Culture 2009. Retrieved from
20
21 <https://escholarship.org/uc/item/01r0k9br>
22
23

24
25
26
27 Bowker, G., Baker, K., Millerand, F., and Ribes, D. (2010) 'Toward Information Infrastructure
28 Studies: Ways of Knowing in a Networked Environment' in Hunsinger, J., Klastrup, L. and
29 Allen, M. (Eds) *International Handbook of Internet Research* New York, Springer.
30
31
32

33
34
35
36 Bowker, G. (2014) 'The Infrastructural Imagination' in Mongili, A. and Pellegrino, G. (Eds)
37 *Information Infrastructure(s): Boundaries, Ecologies, Multiplicity*. Newcastle Upon Tyne, Cambridge
38
39 Scholars Publishing.
40
41
42

43
44
45 Brügger, N., Milligan, I., Ben-David, A., Gebeil, S., Nanni, F., Rogers, R., Turkel, W.J., Weber,
46 M.S. and Webster, P., 2019. 'Internet histories and computational methods: a "round-doc"
47 discussion.' *Internet histories*, 3(3-4), pp.202-222.
48
49
50

51
52
53 Floridi, L. (2009) 'Web 2.0 vs. the Semantic Web: a philosophical assessment' in *Episteme* 6(1),
54
55 pp.25-37.
56
57
58
59
60

- 1
2
3 Ford, H. and Graham, M. (2016) 'Semantic Cities: Coded Geopolitics and the Rise of the
4 Semantic Web in Kitchin, R. and Perng, S-Y. *Code and the City*, London, Routledge.
5
6
7
8 Gillespie, T. (2010). 'The politics of "platforms."' *New Media & Society*, 12(3), pp.347–364.
9
10
11 Gray, J. Gerlitz, C. and Bounegru, L. (2018) 'Data Infrastructure Literacy' *Big Data and Society*,
12 5(2).
13
14
15
16 Halford, S., Pope, C. and Weal, M. (2013) 'Digital Futures? Sociological Challenges and
17 Opportunities in the Emergent Semantic Web'. *Sociology* 47(1), pp. 173-189
18
19
20
21 Halford, S., Weal, M., Tinati, R., Carr, L. and Pope, C. (2018) 'Understanding the production and
22 circulation of social media data: Towards methodological principles and praxis'. *New Media and*
23 *Society* 20 (9), pp.3341-3358
24
25
26
27
28
29 Halpin, H. and Monnin, A. (2016) 'The decentralization of knowledge: How Carnap and
30 Heidegger influenced the Web' *First Monday* 21(12)
31
32
33
34 Helmond, A. (2015). 'The Platformization of the Web: Making Web Data Platform Ready'. *Social*
35 *Media + Society*, 1(2).
36
37
38
39
40 Holmes R (2013) 'From inside walled gardens, social networks are suffocating the internet as we
41 know it.' *Fast Company*, 9 August. Available at:
42 [http://www.fastcompany.com/3015418/frominside-walled-gardens-social-networks-are-](http://www.fastcompany.com/3015418/frominside-walled-gardens-social-networks-are-suffocating-the-internet-as-we-know-it)
43 [suffocating-the-internet-as-we-know-it](http://www.fastcompany.com/3015418/frominside-walled-gardens-social-networks-are-suffocating-the-internet-as-we-know-it)
44
45
46
47
48
49 Jackson, S., Edwards, P., Bowker, G., and Knobel, C. (2007) 'Understanding Infrastructure:
50 history, heuristics and cyberinfrastructure policy' *First Monday*. 12(6).
51
52
53
54 McCarthy, M. (2017) 'The Semantic Web and its entanglements' *Science, Technology & Society* 22(1)
55 pp.21–37
56
57
58
59
60

- 1
2
3 Mukherjee, R. (2019). 'Jio sparks Disruption 2.0: infrastructural imaginaries and platform
4 ecosystems in 'Digital India.'" *Media, Culture & Society*, 41(2), pp.175–195.
5
6
7
8 Musiani, F. (2015). 'Practice, Plurality, Performativity, and Plumbing: Internet Governance
9 Research Meets Science and Technology Studies.' *Science, Technology, & Human Values*, 40(2),
10 pp.272–286.
11
12
13
14
15 Musiani, F. and Schafer, V. (2019) 'Science and Technology Studies Approaches to Web History'
16 in Brügger, N. and Milligan, I. (Eds) *The SAGE Handbook of Web History* London, Sage.
17
18
19
20
21 O'Reilly, T. (2005). 'What is Web 2.0' (30 September), at
22 <http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html>
23
24
25
26 Poirier, L. (2017) 'A Turn for the Scruffy: an ethnographic study of semantic Web architecture'
27 *WebSci17*, June 25-8, 2017, Troy NY, USA
28
29
30
31 Poirier, L. (2018) 'Making the Web Meaningful: a history of Web semantics' in Brügger, N. and
32 Milligan, I. (Ed) *The Sage Handbook of Web History* London, Sage.
33
34
35
36 Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2018). 'Infrastructure studies meet
37 platform studies in the age of Google and Facebook.' *New Media & Society*, 20(1), pp.293–310.
38
39
40
41 Plantin, J.-C., & Punathambekar, A. (2019). 'Digital media infrastructures: pipes, platforms, and
42 politics'. *Media, Culture & Society*, 41(2), pp.163–174.
43
44
45
46 Ribes, D. and Bowker, G. (2005) 'A Learning Trajectory for Ontology Building' In *Annual*
47 *Knowledge and Organizations Conference*, Long Beach, CA.
48
49
50
51
52 Risso, L. (2018) Harvesting your soul? Cambridge Analytica and Brexit. In: Proceedings of the
53 Brexit Means Brexit?, Mainz. Available at:
54 [http://www.adwmainz.de/fileadmin/user_upload/Brexit-Symposium_Online-](http://www.adwmainz.de/fileadmin/user_upload/Brexit-Symposium_Online-Version.pdf#page=75)
55
56
57
58
59
60

1
2
3 Sandvig, C. (2013) 'The Internet as Infrastructure' in Dutton, W. (Ed) *The Oxford Handbook of*
4
5 *Internet Studies* Oxford, Oxford University Press.

6
7
8 Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., Le Roux, B.,
9
10 Friedman, S., & Miles, A. (2013). 'A New Model of Social Class? Findings from the BBC's Great
11
12 British Class Survey Experiment'. *Sociology*, 47(2), pp.219–250.

13
14
15 Savage, M. (2018) *Social Class in the 21st Century* London, Penguin.

16
17
18 Schubert, C. and Kolb, A. (2019) 'Designing Technology, Developing Theory: Toward a
19
20 Symmetrical Approach' *Science, Technology and Human Values* 46(3) pp.528-554.

21
22
23 Shirkey, C. (2006) 'The Semantic Web, Syllogism and Worldview' Clay Shirkey's Writings about
24
25 the Internet

26
27
28 http://eolo.cps.unizar.es/docencia/doctorado/Articulos/WebSemantica/Semantic%20Web%20-%20Shirky_%20The%20Semantic%20Web,%20Syllogism,%20and%20Worldview.pdf
29
30
31

32
33
34 Star, S. L. (1999) 'The Ethnography of Infrastructure' *American Behavioral Scientist* 43(3), pp.377-
35
36 391.

37
38
39 van Dijck, J., Poell, T. (2013). 'Understanding Social Media Logic.' *Media and Communication*, 1(1),
40
41 pp.2-14.

42
43
44
45 van Heur, B., Leydesdorff, L., & Wyatt, S. (2013). Turning to ontology in STS? Turning to STS
46
47 through 'ontology.' *Social Studies of Science*, 43(3), pp.341–362.

48
49
50
51 Welty, C. and Fikes, R. (2006) 'A Reusable Ontology for Fluents in OWL' in Bennett, B. and
52
53 Fellbaum, C. (Eds) *Formal Ontology in Information Systems* Amsterdam, IOS Press.

54
55
56
57
58
59 Winner, L. (1986) 'Do Artefacts Have Politics?' *Daedalus* 109(1), pp.121-136.
60

1
2
3 Winograd, T., 1975. Frame representations and the declarative/procedural controversy. In
4
5 *Representation and understanding* (pp.185-210). Morgan Kaufmann.

6
7
8 Woolgar, S., & Lezaun, J. (2013). The wrong bin bag: A turn to ontology in science and
9
10 technology studies? *Social Studies of Science*, 43(3), pp.321–340.

11
12
13
14
15 ⁱ July 2022 [https://www.statista.com/statistics/617136/digital-population-worldwide/January 2022](https://www.statista.com/statistics/617136/digital-population-worldwide/January-2022). Accessed 1 July
16 2022.

17 ⁱⁱ The original proposal for the WWW, HTMLized (w3.org), available from:
18 <https://www.w3.org/History/1989/proposal.html>

19
20 ⁱⁱⁱ <https://videos.cern.ch/record/2671957>. Recording of a talk by Tim Berners-Lee at the First International
21 Conference on the World-Wide Web hosted by the European Organization for Nuclear Research (CERN) in 1994.
22 Accessed 1 August 2022.

23
24 ^{iv} We are grateful to the ESRC for supporting this research, grant number ES/M0003809/1

25 ^v Unlike GBCS, this assigns individuals to social class as if it's a fact, rather than a likelihood—epistemologically this
26 is a different way of generating information than latent class analysis, from inductive to deductive.

27 ^{vi} This research received ethical approval from the University of Bristol and the University of Southampton.

28 ^{vii} The term ontology is used in a related but different way in Social Science, and this distinction had long been used
29 in the team as a marker of the complexities of interdisciplinary research. Here we describe what “building an
30 ontology” means in the semantic Web. Later we consider some of the ontological implications of this from a
31 sociological perspective.

32
33 ^{ix} A mapping language to facilitate conversion from structured data (e.g. csv or database tables) into a linked data
34 format.

35 ^x Triples are the basic building blocks of RDF (Resource Description Framework), the standard SW modelling
36 language.

37 ^{xi} The general inferential capabilities afforded by logic-based ontology languages (as opposed to algorithms) for
38 deriving specific entailments.

Biographies

Susan Halford is Professor of Sociology. She was a founding Director of the Web Science Institute at the University of Southampton and is currently co-Director of the ESRC Centre for Sociodigital Futures at the University of Bristol. Her research focuses on the politics and practices of digital data, artifacts and infrastructures, with specific attention to the processes of creating and intervening in emergent sociodigital futures.

Mark Weal is a Professor of Web Science in the Digital Health and Biomedical Engineering Group in Electronics and Computer Science at the University of Southampton. He is an Associate Director of the Web Science Institute and co-director of the LifeGuide project to develop digital public health interventions. His research interests include the application of Semantic Web technologies and information systems, including hypermedia systems and eLearning technologies.

Faranak Hardcastle is a Research Fellow in the CELS-Oxford research group at the Wellcome Centre for Human Genetics. Her research uses insights from Science and Technology Studies and Critical Data and Algorithm Studies to critically engage with the development and application of new healthcare and Web technologies.

Nicholas Gibbins is an Associate Professor in the Web and Internet Science group in the University of Southampton's School of Electronics and Computer Science. His key areas of research are the infrastructure of the Semantic Web, the design and development of knowledge-intensive applications. His teaching at Southampton closely follows his research: Semantic Web, hypertext, databases and artificial intelligence.

Samantha Kanza is a Senior Enterprise Fellow at the University of Southampton. She coordinates the AI 4 Scientific Discovery Network (AI4SD) and the Future Blood Testing Network. Her research applies computer science techniques to the scientific domain, through the use of semantic Web technologies and artificial intelligence. She has worked on interdisciplinary Semantic Web projects in different domains, including agriculture, chemistry, and the social sciences.

Catherine Pope is Professor of Medical Sociology and senior research fellow at Green Templeton College, University of Oxford. She co-led the Doctoral Training Centre for Web Science at the University of Southampton. An expert in qualitative and mixed methods for applied health research, she is a key contributor to developing methods for evidence synthesis. She has published empirical, theoretical, and methodological work for clinical, sociological, policy and practitioner audiences.

1
2
3
4
5
6
7 TECHNICAL-MIDDLE

8
9 (HHIncome = High AND SocialContactNumber < Moderate) - agreed

10
11 OR

12
13 (HHIncome = ModeratelyHigh AND HValue > ModeratelyLow AND HHSavings <
14 VeryHigh AND HHSavings > Moderate AND SocialContactNumber = VeryLow) - agreed
15
16

17 Figure 1. Fragment of description of ontology inference rules derived from GBCS literature.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review