

# The Hidden History of the Like Button: From Decentralized Data to Semantic Enclosure

Harry Halpin 

Social Media + Society  
July-September 2023: 1–13  
© The Author(s) 2023  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/20563051231195542  
journals.sagepub.com/home/sms  


## Abstract

Artificial intelligence relies on the use of semantic technologies to represent the shared world of humanity. The story of how this came to be is exemplified by the use of Semantic Web standards by the Facebook “Like” button. In the case of the “Like” button, a decentralized and open Semantic Web was used to fuel the accumulation of personal data for advertising throughout the entire Web. The advent of the “Like button” was shortly followed by Google’s creation of the Google Knowledge Graph, a private corporate version of the Semantic Web. In fact, every major company in Silicon Valley soon created its own knowledge graph. The Semantic Web was transformed from a democratic project for standardized open knowledge to a project of control, collapsing semantics and erasing the difference between the object qua object and the object as represented in a knowledge graph.

## Keywords

Facebook, Like button, Semantic Web, semantics, standards

## Introduction

The transformation of the world into digital data itself seems to be taken for granted by former Google CEO Eric Schmidt and Cohen (2013) in their book, *The New Digital Age*. This increasing “datafication” of various aspects of the world is precisely one of the most pressing political issues of our age (Van Dijck, 2014), and we argue that one key moment was when obscure Semantic Web standards were used to create the Facebook “Like” button and then corporate knowledge graphs (Berners-Lee et al., 2001). Earlier work in artificial intelligence (AI) was content merely to represent the world with digital computers such that the knowledge representation had a semantic, and thus non-causal, relationship with that which it represented (Smith, 1996).<sup>1</sup> With the rise of large proprietary corporate knowledge graphs at the heart of companies such as Facebook, Google, Amazon, and Apple, this semantic relationship has transformed from mere representation to an apparatus that impinges upon and is even meant to control the wider world outside of the digital.

The transformation of the once open Web into a number of “walled gardens” that trap as much as possible of our social life within proprietary platforms like Facebook and Google is a cautionary tale (Gillespie, 2010): How the evolution of the hypertext Web of documents into the Semantic

Web as envisaged by Tim Berners-Lee et al. (2001), the inventor of the Web, went wrong. It is ultimately a hidden history of how the “Like” button is not just the creation of Facebook but also a consequence of the open standards produced by the World Wide Web Consortium (Berners-Lee & Fischetti, 1999), and how the people that meant to create the Semantic Web as the next step to empower ordinary users ended up creating centralized corporate knowledge graphs monopolized by a small number of firms (Chaudhri et al., 2022). Within this history, there is a larger theoretical lesson about the interrelations of the social sphere and the digital sphere, and how the digital sphere—previously thought of as an innocent mirror that reflected the “real world” of “concepts, people, and places” as put by Tim Berners-Lee (1994)—ended up being a powerful force of social engineering via semantic media.

To begin to understand the “Like” button and the Semantic Web, our argument transverses how the vision of the Semantic Web recapitulated the historic trajectory of

American University of Beirut, Lebanon

### Corresponding Author:

Harry Halpin, American University of Beirut, P.O. Box 11-0236, Riad El Solh 1107 2020, Beirut, Lebanon.  
Email: hhalpin@ibiblio.org



“good-old fashioned artificial intelligence” (Haugeland, 1989). Then we show how Semantic Web technology meant to create a web of open data for the public good got hijacked by Facebook to create a closed silo of social data for advertising purposes via the “Like” button. This in turn led Google to create their own proprietary version of the Semantic Web, the Google Knowledge Graph, via in turn consuming public data from Wikipedia and the Semantic Web-based *schema.org*. Slowly but surely, semantic media begin to control the objects they once represented, thereby eliding the very semantic relationship itself.

## The Semantic Web

The “Web 1.0” is simply a way to connect documents about the world via hypertext links (Berners-Lee & Fischetti, 1999). This theory of the Web as consisting of links between web pages is embodied in the Hyper Text Markup Language (HTML). In this framework, the economy of the Web is defined as one of attention, where this attention is given by the following of links that create “hits” on a web page (Rogers, 2002). This raw material of links and media remains at the level of pure syntax, and so can be input directly to algorithms like Google’s Pagerank (Page et al., 1999), with various techniques from natural language processing and information indexing allowing the efficient retrieval of web pages. Yet there is always semantics behind syntax. One of the key inventors of the Web, Tim Berners-Lee himself, put forward in his keynote speech at the first World Wide Web Conference (W3C) in 1994, that beneath this world of hypertext documents was a world that could also be directly described by the Web, not just indirectly via documents:

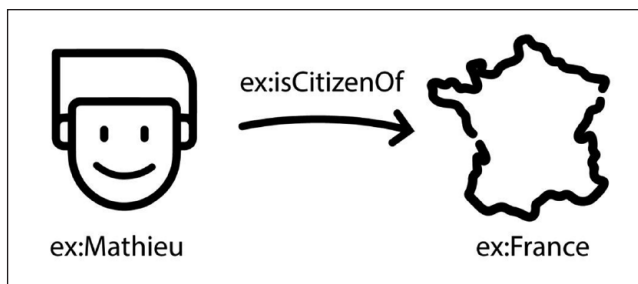
To a computer, then, the Web is a flat, boring world devoid of meaning. This is a pity, as in fact documents on the web describe real objects and imaginary concepts, and give particular relationships between them . . . adding semantics to the Web involves two things: allowing documents which have information in machine-readable forms, and allowing links to be created with relationship values. (Berners-Lee, 1994)

After all, Berners-Lee was a database engineer, not a literary theorist like Ted Nelson, so naturally he was interested in bringing structure to the Web and somehow grasping the wide world outside the Web within schemas. This vision, which seems naïve, belies a totalizing ambition of cosmic proportions: the Semantic Web (Berners-Lee et al., 2001).

Without common protocols that define how machines communicate information and their range of actions in response to information, there would be no interoperability and so no internet as a network of networks. Contra the occasional theoretical confusion, it is hard to imagine any kind of communication network without standards (Galloway, 2004). At the origin of the Web, existing standard bodies such as the Internet Engineering Task Force (IETF) shepherded standards such as Uniform Resource Identifiers (URIs) to completion: the universalizing naming

scheme that lets anyone name a web page in the now familiar syntax given by *https://example.org* (Halpin, 2012). However, the IETF seemed too slow-moving for the Web, as a number of large companies such as Microsoft and Netscape produced browsers with incompatible renderings of HTML. This led to the very real possibility that the Web would syntactically fragment into incompatible corporate fiefdoms (Berners-Lee & Fischetti, 1999). To prevent the fragmentation of the Web and preserve its universality, Tim Berners-Lee founded the W3C at MIT (with eventual offices in France, Japan, and China) as a standards body for the Web dedicated to “leading the Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web.”<sup>2</sup> Unlike the more informal IETF, and leveraging his position as the supposed inventor of the Web, Berners-Lee’s W3C was founded as a “pay to play” organization where companies such as Microsoft (and eventually Google and Facebook) could join via paying membership dues. The membership paid for an international staff that considered themselves a sort of new “Knights of the Roundtable” whose mission was to defend the Web (Berners-Lee & Halpin, 2012). Membership in the W3C was open to any organization whether commercial or non-profit, although governments could only indirectly participate via research grants and could not officially have a seat at the table, unlike the International Standards Organization (ISO). At first glance, the W3C seems to be organized as a strict representative democracy, with each member sending a single member to the Advisory Committee that votes on the creation of new standards and their finalization as W3C Recommendations. Yet formally, the Advisory Committee only advises Berners-Lee, and as such the W3C is in reality more of a constitutional monarchy, as historically decisions in the final instance always are made by Berners-Lee himself and he holds veto power over any proposed W3C standard. A W3C standard is considered open insofar as the copies of the standard are freely available on the Web, and each member of the W3C consortium pledges to license any patents on a royalty-free basis to anyone who implements the standard (Yates & Murphy, 2019). It is precisely this mechanism that enables open innovation on the Web (Ettlinger, 2017) such that anyone can implement a web browser, and so prevents incompatible versions of browsers from destroying the Web.

The W3C was also a vehicle to realize the next step in Berners-Lee et al.’s (2001) plan to evolve the web of documents into the Semantic Web, via a collection of new standards and government-sponsored research projects. The primary standard to accomplish this was the resource description framework (RDF) standard, where any resource could be linked together to any other resource via a typed link (Lassila & Swick, 1999). For example, the fact that “Mathieu is a citizen of France” could be decomposed into a “*isCitizenOf*” link between a URI representing a person (as in their homepage) named Mathieu, and a URI representing France, as given in Figure 1.<sup>3</sup>



**Figure 1.** “Mathieu is a citizen of France” in RDF.

In the mind of Berners-Lee (1998), the Semantic Web was viewed as a fundamentally democratic project, where each person would externalize their knowledge into their own personal knowledge space using RDF and each resource would be eternally named with a URI. In this manner, Berners-Lee and Fischetti (1999) imagined he could transform the web into a giant decentralized database: “Most databases in daily use are relational databases . . . the relationships between the columns are the semantics—the meaning—of the data. These data are ripe for publication as a semantic web page.” Yet when engineers speak of semantics, they leap off the narrow bridge of engineering and fall head first into the murky swamp of semantics, an indistinct inquiry that lies at the heart of the nexus between our everyday cognitive abilities and our submersion in the social (Sellars, 1956).

Although a more full history of the semantics in the Semantic Web is given elsewhere in depth (Halpin, 2012), we can recapitulate the essential ideas briefly. Berners-Lee’s concept of semantics is superficial but gets certain aspects correct, as it involves a relationship between entities in the world and a set of arbitrary symbols given in a syntax, where these symbols are considered representations of entities in the world. These symbols are given an expression in an often formal language where the language defines what syntactic combinations are valid. For example, a language like RDF defines a subject, predicate, and object: *ex:Mathieu ex:isCitizenOf ex:France* (Lassila & Swick, 1999). Unlike HTML which merely defines what syntax can be correctly parsed by a web browser, semantics are part of a network of relationships where certain sentences are true and others are false (or undefined semantically). Furthermore, certain true sentences may infer other true sentences: If Mathieu is a citizen of France, he therefore is European. The trick to correctly defining a semantic relationship is to line up the world and the representations of the world via some underlying logic such that the representation of the world and the semantics of the language map to each other consistently (Smith, 2002). As the world itself is typically inaccessible to a mere computational syntax, a model of the world is built using mathematical structures such that the syntax aligns with the mathematical model to produce predictable (and so computationally automatizable) inferences (Tarski, 1944). The success of the Semantic Web

depended on the creation of a knowledge representation language for the Web, and so the Semantic Web inherits both the successes and failures of previous efforts to create knowledge representation languages in AI.

At the inception of the field of AI, there was no widespread consensus on semantics. John McCarthy (1959), the co-founder of AI, believed that intelligence could be captured by the semantics of logic. McCarthy championed knowledge representation languages with classical predicate first-order logic as put forward by Frege (1972), as first-order predicate logic already had a well-defined semantics (Hayes, 1977). In what appeared to be a revival of the program of the Vienna Circle to unify all human knowledge using logic (Halpin & Monnin, 2016), arch-logician Pat Hayes’ (1979) “Naïve Physics Manifesto” even pushed for the formalization of common-sense physical reasoning using first-order logic. This program was doomed from the start; the impoverished models of first-order logic could not deal with the complexity of the real world, ranging from default reasoning to temporal change (McCarthy & Hayes, 1969). To make matters even more embarrassing, even when first-order logic worked to describe the real world, the inferences were found to be often trivial or irrelevant (McDermott, 1987). So a number of other alternative knowledge representation languages arose, although it could be argued they could mostly be reduced to predicate logic (Hayes, 1977). One prominent challenge was the concept of semantic networks, a term originally used to describe a knowledge representation language for the earliest machine-translation systems (Richens, 1956). Semantic networks are “a graphic notation for representing knowledge in patterns of interconnected nodes and arcs” (Sowa, 1987). They are not new but as ancient as classical propositional logic, as they were first informally used in the explanation of Aristotelian categories by Porphyry (Sowa, 1987). The particular approach of semantic networks was given some credibility by the fact that often when attempting to make diagrams of knowledge, humans start by drawing circles connected by lines, with each component labeled with some human-readable description. This exact formulation was re-invented by Berners-Lee and the W3C with RDF, except rather than using human-readable descriptions in natural language, the nodes and arcs are labeled by URIs that can be read by a web browser (Lassila & Swick, 1999).

Although major companies such as Microsoft and Netscape refused to support the Semantic Web effort, under the guidance of Berners-Lee the W3C continued to essentially commit to a research program by prematurely standardizing the Semantic Web, supported by mostly universities and research grants. With Dan Brickley of the W3C, the former lieutenant of Lenat’s Cyc project to create programs with common-sense using a mixture of semantic nets and first-order logic (Lenat et al., 1990), R.V. Guha created a simple schema language to define different types of entities

in RDF (Brickley & Guha, 2004). Yet those who do not learn from history are doomed to repeat it; the first version of RDF was put forward without any kind of formal semantics, and so found itself prone to error and incorrect inferences (Lassila & Swick, 1999). Hayes (1977, 2004), the staunch defender of logic as a foundation for semantics, created a model-theoretic semantics for RDF to allow RDF statements to make basic inferences. This was quite the momentous moment for logic on the Web, as it was the first logical system that operated at the web scale. Shortly thereafter, an “ontology” language called the Web Ontology Language (curiously abbreviated as OWL), based on the previously obscure field of description logic, was added to allow more complex classes (called “vocabularies” and sometimes “ontologies” despite the philosophical confusion engendered by the usage of that term in the academic logic community), which in turn enabled a limited yet computationally efficient number of inferences (Patel-Schneider et al., 2004). Note that as new information could at any point be added to the Web, RDF did not use standard first-order logic, with its universal statements and the law of the excluded middle. Instead, RDF found itself based on a much simpler “open world” semantics, where things could be proven true but nothing could be proven false as new logical statements could always be added at any time, throwing out much of the semantic machinery of first-order predicate logic.

Despite the immense effort put in a host of complex W3C standards, the decentralized Semantic Web of data (sometimes rebranded as “Linked Data”) never reached widespread adoption. One of its fatal problems was that of linking between RDF statements. On the hypertext Web, linking between web pages is straightforward and results in a clear action by a user when they “click” on a link. On the Semantic Web, links between entities give definitive relationships and enable semantics-driven inferences. The clearest form of linking is to simply assert that two entities are the same, as in the famous Fregean example that Phosphorus (the morning star) and Hesperus (the evening star) actually share the same referent, a single star. Even in semantic networks in AI, there was vast disagreement between researchers on what a “link” actually meant (Woods, 1975). Again, it should come as no surprise that when humans tried to use the Semantic Web, they could also not come to an agreement on linking, terms, or even equality between Semantic Web entities using the logical language of Semantic Web itself (Halpin et al., 2010). This sheer lack of interoperability on the level of shared meaning rendered the decentralization of the Semantic Web difficult, if not impossible. As the entire plan of the Semantic Web was that ontologies would be linked together to create a “bottom-up” space of knowledge, this appeared to be a fatal flaw. After all, the Semantic Web is a problem of social construction as much as a technical design. Even more basic technical problems seemed to doom the Semantic Web: Ordinary people themselves ended up being remarkably poor

knowledge engineers, creating a large amount of relatively broken links between their ontologies (Halpin et al., 2010).

Worse, URIs that were meant to be used for knowledge representation on the Semantic Web tended to go down over time, and so the idea of the Semantic Web as a “Web of Linked Data” ended up not working as the URIs themselves used in RDF would expire due to people forgetting to pay domain name renewal registration fees. As a last-ditch effort to save the Semantic Web, the W3C began working on adding RDF annotations to HTML via W3C RDFa (Adida et al., 2008). However, webmasters found the obtuse world of formal semantics too complex and RDFa suffered from a lack of real-world deployment. Little did the Semantic Web advocates suspect that their first large deployment would be by Facebook.

## How Semantics Empowered the Like Economy

One clear difference between the original Web and the Social Web is that the Web was originally composed of open standardized protocols for documents, yet today the lifeblood of innovation on the Web is closed proprietary algorithms running over secret semantic graphs based on social relationships. Rather than being the result of a “bottom-up” process of social innovation where ordinary people create web pages and user-generated media to increase their autonomy, at the present moment the Web is primarily seen by corporations as a massive data set to feed machine learning algorithms for ever more efficient forms of behavioral prediction, including advertising. One key moment in the evolution could be considered to be the invention of behavioral advertising due to cookie-based user tracking (Lerner et al., 2016). However, data produced by cookie-based tracking is non-conceptual, insofar as it provides only a statistical trace of behavior whose value is realized primarily in a sort of aggregate form (Cussins, 1990). There are no distinct concepts or representations in cookie-based tracking, only statistical probabilities attached to identifiers. To take a straightforward example, a cookie cannot reliably gather data about a single individual beyond their web-surfing habits, and it is fairly trivial for users to either intentionally or even accidentally restrict their tracking across browsers and even origins (i.e., individual websites) via switching browsers or removing cookies (Ishtiaq et al., 2017). A web cookie only tracks a user’s behavior across the Web; it cannot tell in detail whether or not a particular user has a certain friend, who their parents are, their address, political affiliation, sexual orientation, what their bank account balance is, and other more discrete forms of information (except via disturbingly accurate but always error-prone statistical inference). On a methodological level, statistical inference may not be enough: Even if a prediction algorithm is 90% certain someone has a particular address, there is a 10% chance that the address is wrong.

This is how Facebook discovered the Achilles' Heel of Google's statistical-based advertising network. By transforming the Web into a user-centric social platform, Facebook could get users to type in their own names, addresses, gender, friends, and even desires (via the "Like" button), so Facebook could overcome the vagaries inherent in statistical approximation and present an even more accurate advertising network than Google (Dimova et al., 2022). With semantics, the elements of a digital model of the world correspond to various discrete entities and events in the world itself. This relationship between purely mechanical syntax, as given by symbols, and their "real-world" referents, is semantics.<sup>4</sup> The key to the evolution of semantics on the Web is the hidden history of the Facebook "Like" button as a recuperation of the Semantic Web. It is precisely this semantic component of the "Like" button that has been neglected by previous analysis of what has been termed the "Like" economy (Gerlitz & Helmond, 2013). The "Like" economy is an incarnation of what Zuboff calls "surveillance" capitalism, a form of capitalism based on the capture of data and attention (Zuboff, 2019). Contra Zuboff, rather than consider surveillance as a mutation inside of capitalism, the evolution of capitalism toward surveillance and control is a consequence of the tendency within capitalism itself to increase the accumulation of capital via the enclosure of the commons, thereby transforming previously non-capitalist spaces into sites of capital accumulation (Federici, 2018). Our very social relationships are simply the latest frontier of the accumulation of capital (Luxemburg, 1913). This enclosure of the commons on the level of data in general and social data, in particular, can be considered the latest stage in the enclosure of the internet (Meinrath et al., 2011). To appreciate the remarkable twisting of the Semantic Web and personal identity started by Facebook, one should start at the very beginning of the Web: The first form of identity on the original Web was the humble homepage, an often jumbled collection of links, photos, and other text. So it should come as no surprise that the first vocabulary to use the Semantic Web was the eponymous "Friend of a Friend" (FOAF) informal semantic Web vocabulary in 2000 that let anyone encode their social network using RDF as a list of friends with URIs.<sup>5</sup> Invented by Dan Brickley and Libby Miller, soon every W3C staff member and Semantic Web programmer had their own "machine-readable" FOAF page in addition to their homepage. As a FOAF social networking site that put literally any URI in a list of friends, FOAF formed a decentralized social network. A social network as data soon started to be called a social graph by programmers (Breslin & Decker, 2007).

While usage stalled as creating a FOAF file appeared too difficult for many users, FOAF came back to the limelight with the shutdown in 2005 of Orkut, a popular social networking site of the emergent social media sphere at the time (Sreberny & Khiabany, 2010). The Google-owned Orkut social networking site was dominant in the Global South for a short period of time before Facebook and so

the repercussions of the shutdown were felt far and wide. The idea occurred to some social networking advocates that people should be able to "back up" their social network by exporting it from Orkut to FOAF so that the semantics of the social network would be portable between different services using FOAF. For the growing "hacker" community taking shape around open source software and the internet, social network and profile data were literally part of the self, so if they were arbitrarily destroyed, then it was equivalent to harming the integrity of the body, although in this case the integrity stretched into the digital sphere. This became known as the "data portability" movement shortly thereafter (Engels, 2016).

The first to enunciate this radical new view of the importance of digital social networking was none other than Brad Fitzpatrick, the founder of one of the earliest and most popular blogging platforms, LiveJournal. In Fitzpatrick, 2007 essay, "Thoughts on the Social Graph," he tried to answer the provocative question: Who owns the social graph? Fitzpatrick foresaw the rise of Facebook, but he noted that there was "a lot of hesitation" about Facebook, as "a centralized 'owner' of the social graph is bad for the internet" (Fitzpatrick, 2007). Fitzpatrick argued that the graph should not be owned by anyone, but that the social graph was a collective good: In terms of the social graph, it is important "that any one of these sites shouldn't own it; nobody/everybody should. It should just exist," and then Fitzpatrick envisioned a non-profit hosting open source software that "collects, merges, and redistributes the graphs from all other social network sites into one global aggregated graph . . . then made available to other sites" where anyone could also host the software (Fitzpatrick, 2007). He also imagined that the identity of a user on the site could also be managed by standards like OpenID, where a user could seamlessly share their data across different sites, log in with an identity from any site, and export their data to any site (Recordon & Reed, 2006). This was not idle chatter, as Fitzpatrick's collaborator and editor on 'Thoughts on the Social Graph' and OpenID proponent David Recordon went off to join social networking site SixApart to implement this decentralized vision, and LiveJournal itself implemented the export of its social networking data via FOAF.

However, creating a web page was simply too hard for most people, and while there was a flurry of social networking sites from SixApart to Myspace, soon a dominant social networking website emerged that made it trivial to create an identity on the Web: Facebook. Tim Berners-Lee had imagined that the Web could be easily updated and edited like Wikipedia and even Facebook yet his "read-write" Web tooling never took off. Instead, what Facebook allowed appeared empowering: Letting ordinary people easily update and edit a homepage, including a curated list of friends. Unlike FOAF and other visions of the social graph, Facebook also gave its users some modicum of privacy, although of course not from Facebook itself; users could hide their posts from the Web at

```

<html prefix="og: https://ogp.me/ns/">
<head>
<title>À bout de souffle (1960)</title>
<meta property="og:title" content="en_GB"
content="Breathless"/>
<meta property="og:type"
content="video:movie" />
<meta property="video:director"
...
</head>
...
</html>

```

**Figure 2.** “Mathieu likes *Breathless* by Jean-Luc Godard” with the “Like” Button inside HTML as OGP RDFa.

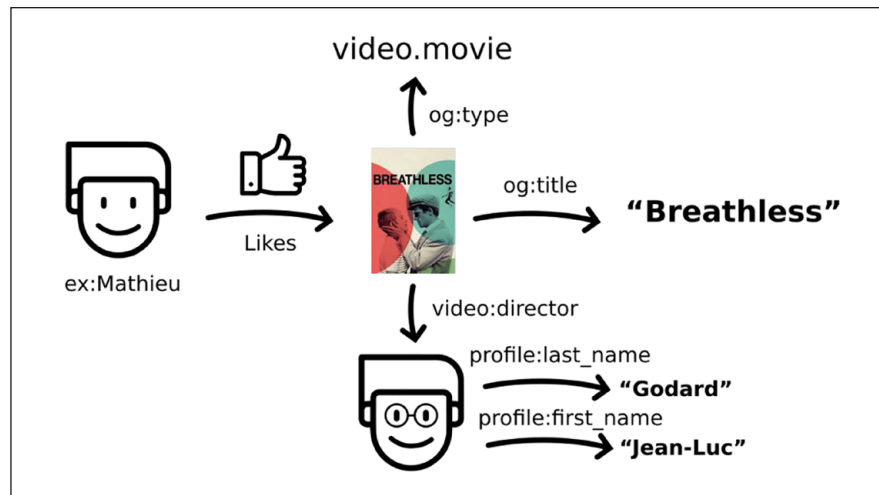
large but reveal them to a curated group of friends. As Facebook grew nearly exponentially, there was a real threat that Facebook would take over the Web. Yet at the very moment that Facebook seemed to be killing the Web, Facebook joined the W3C. The reasons why a closed “walled garden” such as Facebook would join W3C appeared to be that Facebook neither controlled a browser nor one of the rising smartphone platforms, so it made sense that Facebook would support the standardization of HTML and mobile. Strategically, it made no sense for Facebook to “open” its social graph and allow other companies, or even their own users, access to their social graph. Facebook also hired Recordon of “Thoughts on the Social Graph” fame to manage their relationship with the W3C.

A “Like” button was launched by the video-hosting site Vimeo in 2005. Another “Like” button was released in 2007 by Bret Taylor of FriendFeed, a social media aggregation platform. However, these early “Like” buttons could only be used internally on the platform. The feature was invented for Facebook in 2007 by an internal team of engineers, including Leah Pearlman and Justin Rosenstein, and announced in 2009 (Oremus, 2022). The initial concept in terms of advertising was simple: When a user hit the “Like” button, the same content was advertised to their friends. Surprisingly for the W3C, Facebook seemed interested in the Semantic Web, including near-dead standards such as RDFa that had been removed from HTML by browser vendors like Google (Adida et al., 2008). David Recordon even invited W3C staff member Harry Halpin and Pat Hayes (2004), creator of the formal semantics of RDF, to Facebook for a private discussion. The sudden interest of Facebook in the W3C and the Semantic Web was a source of mystery until David Recordon revealed the Open Graph Protocol (OGP).<sup>6</sup> The OGP allowed Facebook’s new “Like” button to extend outside the walled garden of Facebook and into the wider Web. The OGP would allow anyone to easily markup products, movies, and other items with a simplified set of RDFa-style relations in the metadata of *any* web page.

The OGP was neither open nor a protocol, but instead was a way of marking up the declaration of a “like” in a web page. By putting the semantics inside the web page rather than in a separate file or inline in HTML as hoped for by the Semantic Web community, OGP eliminated errors by developers who could then simply cut and paste a “Like” button into a web page, as shown in Figure 3. Facebook also eliminated the open world aspect of the Semantic Web as everyone was forced to use the centralized and closed formal ontology of the (rather ironically-named) OGP. Still, OGP made the Semantic Web actually easy to use by users and developers. This ease of use fed the business model of Facebook, as each “Like” button reported back to Facebook each time a logged-in Facebook user visited a web page with the embedded button so that this tracking information could be used for targeted advertising by Facebook itself.

Within months of the launch of the OGP, the Facebook “Like” button covered the web with RDFa, making the Semantic Web a ubiquitous success, although in a form that Tim Berners-Lee could have never imagined. Rather than formalize all knowledge in logic, as envisioned by Hayes during the heyday of AI, Facebook created a scheme to use the Semantic Web to formalize commercial products on the Web in a privatized and closed graph inside of Facebook. The OGP RDFa that is hosted at the top of a website to insert a “Like” button for the film *Breathless* by Jean-Luc Godard is given in Figure 2, while the extracted RDF from the self-same RDFa is given in Figure 3. However, the problem should be apparent if one contrasts the use of the “Like” button with the Semantic Web RDF example given in Figure 1. Namely, the person—Mathieu in our example—doing the clicking of the “Like” button is identified by a cookie, rather than a public URI, and so is hidden in Facebook’s databases. Similarly, the fact Mathieu clicked the “Like” button is an action shipped straight to Facebook by the cookie and Javascript, and not exposed as open data for anyone else, including Mathieu himself. By the remarkable move of combining cookies with Facebook profiles via Javascript with RDF, Facebook used open data to create a closed world.

This was only the first step. To make the “Like” button work, the user had to be logged into Facebook, so the goal became to have the user logged in all the time. The second step of Facebook was to have other websites go to Facebook to access social data. This could be done with the user’s permission if the user used Facebook Connect.<sup>7</sup> Literally building on top of the single sign-on solution OpenID created by the self-sovereign identity movement (Recordon & Reed, 2006), a user upon entering a new social website could simply automatically login to Facebook to bring whatever personal data automatically to the site—including their list of friends. This was a radically successful coup to control not only the social graph but the identities of people across the entire Web. What Facebook began was nothing more than the primitive accumulation of social relationships; the OGP



**Figure 3.** RDF of “Mathieu likes *Breathless* by Jean-Luc Godard” with the “Like” button.

in combination with Facebook’s profile data reified social relationships and their related affects to create a giant database for advertising purposes. This was also a coup against Google’s dominant advertising network (Dimova et al., 2022): The “Like” Button and Facebook Connect allowed users to be tracked across not just Facebook, but the entire Web, with the smallest details of their buying habits, affects, and favorite websites being exploited by Facebook. From a capitalist perspective, Facebook itself created a monopoly on these data as a platform (Gillespie, 2010), as while the OGP’s “Like” button was open for anyone to use in HTML, the tracking of each user with cookies was a secret held by Facebook. This means that only Facebook could combine the affects of these users with their social relationships, and these emotional relationships were stored inside the closed database of Facebook. There were feeble attempts to create an open-source “Like” button, the OpenLike standard, but without the huge existing user base of Facebook, any alternative to the “Like” button failed to gain traction. Thus the Semantic Web ended up completely reversing control away from the user; the most important part of the data, *who* likes what is controlled only by Facebook.

Prior analysis of the “Like” button failed to take into account the undeniable importance of the Semantic Web, and semantics in general, in the creation of the “Like” button (Gerlitz & Helmond, 2013). Far more than a clever détournement of open standards, the “Like” button ended up being central to Facebook’s advertising network, despite initial skepticism from Mark Zuckerberg himself that the act of “liking” would distract from the creation of social content (Gerlitz & Helmond, 2013). The capture of these social relationships operated via the formalization of social relations, and then the relationships of those individuals to objects, as a mathematical graph. This naturally leads to the possibility of controlling these relationships via the subtle “nudging” of

behavior, such as displaying particular ads in Facebook’s feed or web-scale ad networks (Apte, 2020). Despite the idea that “data is the new oil,” in terms of the dominance of platforms, it is actually only one type of data—personal data—that is the true fuel of advertising (Hirsch, 2013).

As explained earlier, Facebook leaped ahead of Google as while Google uses data to index, Google can’t tell who likes what on the granularity of the individual and their social network with the certainty of Facebook. Ironically, the Semantic Web’s greatest victory, the “Like” button, is precisely what enabled the transition from the “Wild West” open Web of documents in the 1990s to a Web controlled by closed platforms, where the social data was in the hands of Facebook rather than users (Van Dijck, 2014). By capturing the heart of social relationships, including our relationship with things, Facebook got a jump-start in fighting Google’s power of digital advertising. However, social relationships were only the beginning of the accumulation of the lifeworld by capitalism.

### Knowledge Graphs as Corporate Platforms

It should then come as no surprise that other companies followed after Facebook in the race for capitalizing on the Semantic Web. The accumulation and exploitation of social relationships was undergirded by the semantic structure of the “Like” button’s OGP, but what if anything could be the subject of semantic enclosure? A semantic enclosure is an enclosure in which the meaning of a term is captured and integrated into capital. Indeed, the question of using semantics in HTML was not invented by the “Like” button, but by the little-known RDFa standard that allowed the embedding of Semantic Web statements directly inside of HTML. Peter Mika, a Semantic Web researcher at Yahoo! Research,

developed the “SearchMonkey” tool that allowed web page owners to embed any vocabulary in an HTML web page using RDFa and then have these data extracted for display in Yahoo!’s search page (Haas et al., 2011). In addition, SearchMonkey extracted RDF from microformats, a competing technology to RDFa that embedded a finite number of vocabularies into span tags in HTML (Khare & Çelik, 2006). However, the extracted RDF was often itself error-prone; the larger problem was a lack of a common ontology. While the Semantic Web envisioned ordinary people formalizing their knowledge and coming to common agreement via explicit social consensus, globally only hundreds of people were capable of formalizing knowledge with the tooling available at the time. Rather than consensus, if the Semantic Web had reached widespread usage, it is more likely there would have been ontological conflict based on power relations. In fact, this has already been exemplified by the conflict over the status in Wikipedia of Jerusalem in the occupied territories (Ford & Graham, 2016).

A small aside is necessary to clarify the terminology. Entities tend to be formalized in what is called a knowledge base, where entities have properties with values, such as the example of “Mathieu is a citizen of France” given in Figure 1. Less pretentiously called a “vocabulary” or a “schema” of some sort, an ontology is a classification of a set of entities such that certain abstract classes inherit properties and restrictions: From Figure 3 one can deduce that “Every movie has a director.” In particular, most ontologies classify entities by making assertions about the entities based on their classification via a syntactic—and thus, computationally implementable—inference procedure. Entities and ontologies themselves can be linked together, and so form a knowledge graph. The Semantic Web is a set of standards for formal ontologies (OWL and RDF Schema) and entities (RDF) that by virtue of using URIs as identifiers for the names of entities and elements in the ontology thereby allows the naturally linking data into a knowledge graph. However, this is not to say that all knowledge graphs must use the relatively obtuse Semantic Web standards. On the contrary, the term “knowledge graph” came into widespread usage with the release of Oracle’s property graphs, which do not use URIs or Semantic Web standards, but instead just added key-value pairs to relationships and entities in a database (Das et al., 2014). Far more popular than the Semantic Web in commercial environments, property graphs allow efficient querying and are by default captured by vendors in proprietary formats.

Prior to the advent of the “Like” button, Google had handily ignored the Semantic Web. There seemed to be no benefit to Google in investing in the technology, and proponents of the Semantic Web seemed like a fringe cult that had infected the W3C with an idiosyncratic vision of open data. However, the mild uptake of SearchMonkey and the spectacular success of the “Like” button did not go unnoticed in the Googleplex. What had always been missing from the

Semantic Web was a driver of adoption: There was no reason anyone would create semantics for their data, and no applications used the data. The countervailing trend toward the Semantic Web was a general movement toward Javascript and JavaScript Object Notation (JSON)-based application programming interfaces (APIs). As SearchMonkey showed, if developers believed—even if it was not true—that adding semantics to a web page would increase the number of hits that the web page would receive, then developers would add semantic markup like RDFa only for reasons of search engine optimization. While Yahoo! failed to take advantage of SearchMonkey,

Google released its own competitor: Google Rich Snippets.<sup>8</sup> Google Rich Snippets initially allowed a small number of ontologies in RDFa to be displayed as structured data in the Google search results. Within months, Rich Snippets soon included ontologies of people, reviews, events, and even the complexity of food recipes. For example, La Louver could mark its opening times and hours on its web page via a standardized semantic ontology rather than purely via text. Rather than forcing a user to click through a web page to discover its opening hours, the search “When is La Louvre open?” would return the precise opening hours of La Louver on the result page of Google; the user did not even have to click on La Louver’s web page. This real-world search optimization started driving massive adoption of RDF via Google Rich Snippets; there was a world waiting to be captured via semantic media.

Google decided to create its own centralized ontology called *schema.org*. Google also had previously hired the author of RDF Schema, R.V. Guha (Brickley & Guha, 2004). Google then hired Tim Berners-Lee’s former right-hand man and co-lead of RDF Schema, Dan Brickley. In 2011, they launched *schema.org* to build actual social consensus on a single large schema for Rich Snippets (Guha et al., 2003). In the beginning, *schema.org* was considered a semantic peace treaty between search engines, as other major search engines such as Microsoft Bing and Yandex had begun explorations in semantic search and so needed to standardize ontologies for their search engines. Rather than fracture the world of semantic markup into incompatible ontologies, given the market dominance of Google, it made more commercial sense for other search engines to follow the lead of Google. Later *schema.org* accepted, with approval from Google, submissions of new ontologies from the public.

The precise manner in which semantics was to be added to the Web by *schema.org* was the subject of debate. As *schema.org* usage grew and the size of its ontologies also took off, the need to de-facto standardize how to add semantics to web pages became crucial. Similar to Facebook’s “Like” button, at first *schema.org* used a subset of RDFa, but it eventually found the complexity of RDFa caused too many errors by web developers. Also, via focusing on many overly complex Semantic Web standards, the W3C had lost control of HTML. HTML was now controlled by an informal group

of browser developers called WHATWG, dominated by Google. The WHATWG produced an even simpler standard for semantic annotations to web pages called microdata, which took the simplicity of microformats and generalized it for arbitrary data.<sup>9</sup> Struggling for relevance, the W3C then aligned its new standard “RDFa Lite” with microdata, essentially unifying how web pages could embed semantic media (Sporny, 2015). Even in the case of the Semantic Web, the standards eventually aligned with the needs of Google.

While Facebook was dependent on web developers manually adding the “Like” button, for Google humans were brought in as mere useful appendages to AI. The Semantic Web of the W3C envisioned small ontologies “loosely connected together” in the same manner that the Web itself was formed in the 1990s (Weinberger, 2008). This vision of formal ontologies never reached critical mass, with only a few small ontologies such as FOAF gaining any traction and only highly structured domains such as biomedical data being adaptable to fairly rigid schemas (Halpin et al., 2010). In contrast, the phenomenal growth of Wikipedia organically was creating a sort of universal and constantly updated database. Semantic Web researchers began the massive undertaking of transforming the text in the structured parts of Wikipedia into a massive Semantic Web ontology called DBPedia (Bizer et al., 2009). There were also startups moving into semantics. Co-founded by a student of Marvin Minsky, Danny Hillis, the startup Metaweb produced a large-scale ontology called Freebase with over 125,000 entities and 4,000 classes.

Unlike DBPedia, Freebase did not use Semantic Web standards and featured a collaborative interface for human editing of ontologies. At the same time, as Yahoo! haplessly shut down SearchMonkey, Google acquired Freebase in 2010. To separate their corporate efforts from the wider Semantic Web effort, Google relabelled their effort the Google Knowledge Graph in 2012, and likely based their technology on Freebase and an ingested DBPedia, although the exact source of the data of the Knowledge Graph is unknown. It seems likely that over time more and more of the growth of the Google Knowledge Graph is from machine learning and other sophisticated natural language processing techniques and less from *schema.org*. Although Google lacked the ability to explicitly link to humans and their social networks like Facebook, Google created an inhumanly vast database that attempted to formalize all existence, all entities, and all possible relationships between them.

Given the astounding success of the Google Knowledge Graph, soon each company in Silicon Valley was building its own closed knowledge graph. Each company built a knowledge graph customized to its own needs: Amazon built the Amazon Product Graph specialized in determining the number of lots of products (Karamanolakis et al., 2020) while Microsoft started its own knowledge graph about the personal data shared in Microsoft Office (Noy et al., 2019). Within each company, the knowledge graph ends up being a

semantic “glue” that holds together the vast amounts of syntactic data across myriad databases throughout their enterprise. Each of these corporate databases is ultimately about real people as customers, affects as desire for products, places as customer locations, and so forth. While Freebase and Google’s *schema.org* were to a large extent manually built by humans, soon every single Silicon Valley platform was heavily investing in AI techniques to extract semantics from web pages and other sources of data for consumption by their very own knowledge graph. In particular, there was a focus on named-entity recognition and natural language processing to map variegated names to entity nodes and all sorts of verbs and attributes to relations between nodes (Wang et al., 2020). One of the most mysterious companies that seem to have arrived late to knowledge graph construction, Apple seems to have built its knowledge graph entirely by combining its closed data with vast amounts of machine learning (Ilyas et al., 2022). The Semantic Web had publicly and manually constructed semantics that separated mere entities from clear-cut classes, but this effort transformed into competitive closed knowledge graphs where the vast amount of data was automatically discovered entities extracted by AI. In this way, the open world envisioned by the Semantic Web transformed into the closed world of corporate knowledge graphs, where any social conflict over ontological categories has been hidden from the view of the public. These knowledge graphs are the secret power behind the algorithms of search and advertising that dominate the internet today.

## Conclusion: The End(s) of Semantics

The Social Web is ultimately the accumulation of our social relationships into the Web, while in contrast, the Semantic Web incorporates our relationship to things into the Web. However, the very notion of semantics is a normative relationship that defines the kinds of inferences and relationships that are possible. What the Semantic Web did is, via the path of formal model theory, made these possible relationships explicit. It does seem ultimately that the lack of attention paid to semantics by previous analyses of the Social Web (Song, 2010) has ignored the fact that digital representations of people and things on the Web are increasingly indistinguishable from the people and things in the analog world. A person’s self-worth may be determined by the amount of “likes” they receive and a restaurant’s future may be determined by the amount of stars in a rating in Google. This at first seems to be nothing less than an act of ontological violence, where the complexity and particularity of our relationships are subsumed by a certain digital logic. In other words, we find ourselves diverging from the primary scientific program of AI that was inherited by the Semantic Web, to describe the world, to a much more radical program: to control the world. After all, advertising exists to create purchase intent, and AI techniques based on semantic markup can be

also used to create any intent, such as voting intent. David Recordon, after working on the semantics of the “Like” button, ended up moving to the White House. Of course, whether advertising or any technology can predictably create actual intent is a matter of debate (Liu-Thompkins, 2019).

There may have been a philosophical flaw at the heart of the project of the Semantic Web that doomed it from the start. Behind the program of knowledge representation, there is a long-forgotten philosophical program returned from the grave: the philosophy of logical positivism pioneered by Carnap (Halpin & Monnin, 2016). This philosophical program of the Vienna Circle, based on the creation of a scientific and logical meta-language to unify human knowledge, was intended for the common good of humanity, although it was later decried as logical positivism where large swathes of human meaning were discarded due to being resistant to formalization. This philosophy was transformed into concrete engineering via AI in the 1970s by John McCarthy, Pat Hayes, and others. Thus a certain logical positivism inherent in the Semantic Web was twisted to lay the secret foundations of the “new digital age” of Silicon Valley, where any part of life was thought to either succumb to digitization or be rendered unimportant (Schmidt & Cohen, 2013). To explain how the Web has become the locus of sociality and meaning today, there must have been some hidden moment when philosophical questions transformed into engineering questions so that the social could become subsumed into the digital. It is precisely this transversal between engineering and philosophy that needs to be explained so that we can understand what is at stake in corporate knowledge graphs. Our hypothesis is that the turning point between the 19th century—the era when these questions of semantics as meaning were delegated from philosophy to mathematical logic—and today was the formation of the field of AI: The question of meaning once controlled by semanticists then passed to research engineers (Halpin & Monnin, 2016). We further hypothesize that the moment the question of meaning passed from academic AI researchers to corporate engineers was the meeting between Pat Hayes, the arch-semanticist of the Web, and David Recordon of Facebook, from whence the semantics of the “Like” button developed.

To overcome this impasse, what seems to be necessary is a philosophy of meaning that not just resists formalization—as the formalization of knowledge graphs is arguably effective—but puts knowledge formalization in its proper place via a new foundation for semantics. The closest we have to such an articulation lies in the work of Brian Cantwell Smith’s notion of “reconciliation” between the world and its representation, where the preponderance of the world takes precedence over the instrumentalized representations (Smith, 1996). Cantwell Smith says that “abstraction is a process of going from the ineffably rich microdetails of a particular circumstance to an abstract or categorical characterization of it . . . but what about the reverse?” Cantwell Smith (1996) in turn defines “reconciliation” as coming back into effective

contact with what has been registered, and filling in the details.” While returning to complexity seems to be the right move, what Cantwell Smith misses is that with knowledge graphs—of which the “Like” button can be conceived of only as the most simple type—what is happening is a reconciliation that is harnessed to the dictates of contemporary capitalism.

As opposed to the concept of reconciliation as postulated by Cantwell Smith, which fills in the details and thus aligns the universal with the singular, negative reconciliation is precisely a digital inversion wherein the singular is aligned with the formal. This form of reconciliation does not align in some pseudo-Hegelian manner a particular individual with the universal concept: aligning Napoleon, with all his foibles, with some universal concept of man. Instead, what happens is that the singular individual is forced to align with the digital representation of themselves as formalized by a knowledge graph. It is in this manner that we are all becoming subservient to our profiles, whether on Facebook, Instagram, LinkedIn, or whatever is shown by a Google search. These all-too-digital forms assume an ontotheological force that subsumes what was once termed the analog “real.” In other words, the problem is not the reduction of possible affects, which Gerlitz and Helmond theorize explains the absence of a polemical “Dislike” button (Gerlitz & Helmond, 2013). The root of the problem is that there has been a vast semantic enclosure, and the present semantic collapse between the represented and the representation has transformed these formerly semantic systems of knowledge graphs into non-semantic cybernetic systems of control. The non-conceptual limits of Google’s cookie-driven tracking were disrupted by Facebook’s profile page insofar as Facebook provides the precise personal data missing from Google, and these data were in turn weaponized by the “Like” button into a Web-spanning ad network that disrupted Google’s hegemonic ad network. Yet at least Facebook lets data subjects type their own data into profiles, while in Google the profile is created *ex nihilo*; or to be more precise, the profile is created from everything, as the vast data banks that fuel Google’s Knowledge Graph seem to be a “good enough” approximation of everything. It is precisely this lack of control by the represented that makes the entire operation of these knowledge graphs sinister rather than empowering.

As long as the knowledge graphs are under corporate control, the only way out seems to be to create new spaces of opacity, such as decentralized social networks, via cryptography that can resist the transparency of knowledge graphs. To move from tactics to strategy, the only real long-term resolution is to return to the political vision of the Web as a radically democratic space, as ultimately meaning itself cannot be dictated by stale corporate knowledge that is always falling behind the ever-changing temporal progress of the real. Meaning is something we create together, and semantic enclosure is not inevitable. What is needed is nothing short

of a political and economic revolution that can harness knowledge outside of the narrow dictates of the “Like” economy (Gerlitz & Helmond, 2013). One is reminded of the radical heart of Lyotard’s *Postmodern Condition*, where Lyotard (1984) pushes for what he considers a “quite simple” solution to the possibility of the authoritarian digitization of social life: “Give the public free access to the memory and data banks.” At the present moment, knowledge graphs and statistical models serve as the secret foundations of AI tools such as ChatGPT. Although the times have changed, the imperative of opening up secrets remains the same.

### Acknowledgements

This paper developed from a talk given at the Unlike Us conference in Amsterdam on March 10, 2012, at the invitation of Geert Lovink, and serves as a retrospective on how the decentralization of the Semantic Web ended up serving the cause of centralization over the last decade. The video of the lecture is online.<sup>10</sup>

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Harry Halpin  <https://orcid.org/0000-0003-2143-6965>

### Notes

1. To paraphrase a classic example from Brian Cantwell Smith, a person can destroy a (perhaps digital) picture of the Eiffel Tower, but this has no causal effect on the Eiffel Tower itself (1996).
2. <https://www.w3.org/mission/>
3. To easily write entire URIs, the first part of a URI is given as a “namespace,” such that the namespace is substituted by the whole URI it defines when any action is taken with the RDF statement. In this example, the namespace ex: is given to be <https://example.org/>
4. This is a rather blithe and coarse-grained explanation of semantics in terms of correspondence. For a more comprehensive analysis, the work of Sellars is to be recommended (1956), as well as more recently that of Cantwell Smith (1996).
5. <http://xmlns.com/foaf/0.1/>
6. <https://ogp.me/>
7. <https://developers.facebook.com/blog/post/2008/05/09/announcing-facebook-connect/>
8. <https://developers.google.com/search/blog/2009/05/introducing-rich-snippets>
9. <https://html.spec.whatwg.org/multipage/microdata.html>
10. <https://vimeo.com/775826302>

### References

Adida, B., Birbeck, S., & Mark McCarron Pemberton, S. (2008). *RdFa in XHTML: Syntax and processing*. W3C

Recommendation. <https://www.w3.org/TR/2008/REC-rdfa-syntax-20081014/>

- Apte, D. S. (2020). Explicit autonomy, implicit control: User autonomy in the dichotomous choice architecture of Facebook. *Journal of Creative Communications*, 15(2), 165–176.
- Berners-Lee, T. (1994). *The future of the Web (Keynote at World Wide Web Conference)*. <https://www.w3.org/Talks/WWW94Tim/>
- Berners-Lee, T. (1998). *Cool URIs don't change*. <http://www.w3.org/Provider/Style/URI>
- Berners-Lee, T., & Fischetti, M. (1999). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper.
- Berners-Lee, T., & Halpin, H. (2012). Defend the Web. In J. Bus, M. Crompton, M. Hildebrandt, & G. Metakides (Eds.), *Digital Enlightenment yearbook* (pp. 3–7). IOS Press.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34–43.
- Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia: A crystallization point for the web of data. *Journal of Web Semantics*, 7(3), 154–165.
- Breslin, J., & Decker, S. (2007). The future of social networks on the internet: The need for semantics. *IEEE Internet Computing*, 11(6), 86–90.
- Brickley, D., & Guha, R. V. (2004). *RDF vocabulary description language 1.0: RDF Schema*. W3C Recommendation. <https://www.w3.org/TR/2004/REC-rdf-schema-20040210/>
- Chaudhri, V., Baru, C., Chittar, N., Dong, X., Genesereth, M., Hendler, J., Kalyanpur, A., Lenat, D., Sequeda, J., Vrandečić, D., & Wang, K. (2022). Knowledge graphs: Introduction, history and perspectives. *AI Magazine*, 43(1), 17–29.
- Cussins, A. (1990). The connectionist construction of concepts. In M. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 368–440). Oxford University Press.
- Das, S., Srinivasan, J., Perry, M., Chong, E. I., & Banerjee, J. (2014). A tale of two graphs: Property graphs as RDF in Oracle. In *Extending database technology* (pp. 762–773). <https://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=CF33250BC1D9713C75826D8CA60D3CF3?doi=10.1.1.684.4011&rep=rep1&type=pdf>
- Dimova, Y., Franken, G., Le Pochat, V., Joosen, W., & Desmet, L. (2022). Tracking the evolution of cookie-based tracking on Facebook. In *Proceedings of the 21st workshop on privacy in the electronic society* (pp. 181–196). Association for Computing Machinery.
- Engels, B. (2016). Data portability among online platforms. *Internet Policy Review*, 5, Article 408.
- Ettlinger, N. (2017). Open innovation and its discontents. *Geoforum*, 80:, 61–71.
- Federici, S. (2018). *Re-enchanting the world: Feminism and the politics of the commons*. PM Press.
- Fitzpatrick, B. (2007). *Thoughts on the social graph*. <http://bradfitz.com/social-graph-problem/>
- Ford, H., & Graham, M. (2016). Semantic cities: Coded geopolitics and the rise of the Semantic Web. In R. Kitchin & S.-Y. Perng (Eds.), *Code and the city* (pp. 200–214). Routledge.
- Frege, G. (1972). Concept Script, a formal language of pure thought modelled upon that of arithmetic. In T. Bynum (Ed. and Trans.), *Conceptual notation and related articles* (pp. 101–203). Clarendon.

- Galloway, A. R. (2004). *Protocol: How control exists after decentralization*. MIT Press.
- Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society, 15*(8), 1348–1365.
- Gillespie, T. (2010). The politics of “platforms.” *New Media & Society, 12*(3), 347–364.
- Guha, R., McCool, R., & Miller, E. (2003). Semantic search. In *Proceedings of the international conference on world wide web (WWW)* (pp. 700–709). Association for Computing Machinery.
- Haas, K., Mika, P., Tarjan, P., & Blanco, R. (2011). Enhanced results for web search. In *Proceedings of the ACM SIGIR conference on research and development in information retrieval* (pp. 725–734). Association for Computing Machinery.
- Halpin, H. (2012). *Social semantics: The search for meaning on the Web*. Springer.
- Halpin, H., Hayes, P. J., McCusker, J. P., McGuinness, D. L., & Thompson, H. S. (2010). When owl: sameas isn't the same: An analysis of identity in linked data. In *Proceedings of the International Semantic Web conference* (pp. 305–320). Association for Computing Machinery.
- Halpin, H., & Monnin, A. (2016). The decentralization of knowledge: How Carnap and Heidegger influenced the web. *First Monday, 21*(12), Article 7109.
- Haugeland, J. (1989). *Artificial intelligence: The very idea*. MIT Press.
- Hayes, P. (1977). In defense of logic. In *Proceedings of international joint conference on artificial intelligence* (pp. 559–565). Association for Computing Machinery.
- Hayes, P. (1979). The naive physics manifesto. In D. Michie (Ed.), *Expert systems in the micro-electronic age* (pp. 242–270). Edinburgh University Press.
- Hayes, P. (2004). *RDF semantics*. W3C Recommendation. <https://www.w3.org/TR/2004/REC-rdf-mt-20040210/>
- Hirsch, D. D. (2013). The glass house effect: Big data, the new oil, and the power of analogy. *Minnesota Law Review, 66*, Article 373.
- Ilyas, I. F., Rekatsinas, T., Konda, V., Pound, J., Qi, X., & Soliman, M. (2022). Saga: A platform for continuous construction and serving of knowledge at scale. In *Proceedings of the international conference on management of data* (pp. 2259–2272). Association for Computing Machinery.
- Ishtiaq, A., Abbasi, S. H., Aleem, M., & Islam, M. A. I. (2017). User tracking mechanisms and counter measures. *International Journal of Applied Mathematics Electronics and Computers, 5*(2), 33–40.
- Karamanolakis, G., Ma, J., & Dong, X. L. (2020). Textract: Taxonomy-aware knowledge extraction for thousands of product categories. *Arxiv Preprint Arxiv:200413852*. <https://arxiv.org/abs/2004.13852>
- Khare, R., & Çelik, T. (2006). Microformats: A pragmatic path to the Semantic Web. In *Proceedings of the international conference on the World Wide Web* (pp. 865–866). Association for Computing Machinery.
- Lassila, O., & Swick, R. (1999). *Resource description framework (RDF) model and syntax specification: Deprecated W3C recommendation*, W3C. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>
- Lenat, D. B., Guha, R. V., Pittman, K., Pratt, D., & Shepherd, M. (1990). Cyc: Towards programs with common sense. *Communications of the ACM, 33*(8), 30–49.
- Lerner, A., Simpson, A. K., Kohno, T., & Roesner, F. (2016). Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016. In *Proceedings of USENIX security conference* (pp. 997–1013). Association for Computing Machinery.
- Liu-Thompkins, Y. (2019). A decade of online advertising research: What we learned and what we need to know. *Journal of Advertising, 48*(1), 1–13.
- Luxemburg, R. (1913). *The accumulation of capital*. Routledge.
- Liotard, J. F. (1984). *The postmodern condition: A report on knowledge*. University of Minnesota Press.
- McCarthy, J. (1959). Programs with common-sense. *Nature, 188*, 77–91. <http://www-formal.stanford.edu/jmc/mcc59.html>
- McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). Edinburgh University Press.
- McDermott, D. (1987). A critique of pure reason. *Computational Intelligence, 33*(3), 151–160.
- Meinrath, S. D., Losey, J. W., & Pickard, V. W. (2011). Digital feudalism: Enclosures and erasures from digital rights management to the digital divide. *Advances in Computers, 81*, 237–287.
- Noy, N., Gao, Y., Jain, A., Narayanan, A., Patterson, A., & Taylor, J. (2019). Industry—scale knowledge graphs: Lessons and challenges—Five diverse technology companies show how it's done. *Queue, 17*(2), 48–75.
- Oremus, W. (2022). The curse of the awesome button. In T. Bosch (Ed.) *You are not expected to understand this* (pp. 131–138). Princeton University Press.
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web* [Technical report]. Stanford InfoLab.
- Patel-Schneider, P., Hayes, P., & Horrocks, I. (2004). *OWL Web ontology language abstract syntax and semantics*. W3C Recommendation. <https://www.w3.org/TR/2003/PR-owl-semantics-20031215/>
- Recordon, D., & Reed, D. (2006). OpenID 2.0: A platform for user-centric identity management. In *Proceedings of the ACM workshop on digital identity management* (pp. 11–16). Association for Computing Machinery.
- Richens, R. H. (1956). Preprogramming for mechanical translation. *Mechanical Translation, 3*(1), 20–25.
- Rogers, R. (2002). Operating issue networks on the Web. *Science as Culture, 11*(2), 191–213.
- Schmidt, E., & Cohen, J. (2013). *The new digital age: Reshaping the future of people, nations and business*. Hachette.
- Sellars, W. (1956). Empiricism and the philosophy of mind. *Minnesota Studies in the Philosophy of Science, 1*(19), 253–329.
- Smith, B. C. (1996). *On the origin of objects*. MIT Press.
- Smith, B. C. (2002). The foundations of computing. In M. Scheutz (Ed.), *Computationalism: New directions* (pp. 23–58). MIT Press.
- Song, F. W. (2010). Theorizing Web 2.0: A cultural perspective. *Information, Communication & Society, 13*(2), 249–275.
- Sowa, J. (1987). Semantic networks. In S. Shapiro (Ed.), *Encyclopedia of artificial intelligence* (pp. 1011–1024). John Wiley & Sons.
- Sporny, M. (2015). *RDFa lite 1.1* (2nd ed.). W3C Recommendation. <https://www.w3.org/TR/rdfa-lite/>
- Sreberny, A., & Khiabany, G. (2010). *Blogistan: The internet and politics in Iran*. Bloomsbury.

- Tarski, A. (1944). The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4, 341–375.
- Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208.
- Wang, K., Shen, Z., Huang, C., Wu, C. H., Dong, Y., & Kanakia, A. (2020). Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1), 396–413.
- Weinberger, D. (2008). *Small pieces loosely joined: A unified theory of the Web*. Basic Books.
- Woods, W. (1975). What's in a link: Foundations for semantic networks. In D. G. Bobrow, & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 35–82). Academic Press.
- Yates, J., & Murphy, C. N. (2019). *Engineering rules: Global standard setting since 1880*. John Hopkins University Press.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

### Author Biography

Harry Halpin is a visiting professor at American University of Beirut. He is also the CEO and co-founder of Nym Technologies, a privacy company. Previously, he was a Senior Research Scientist at W3C/MIT in the Technology and Society Domain under Tim Berners-Lee, where he led the standardization of the web cryptography API across all browsers and founded the W3C Social Web Working Group to create decentralized social networking technologies. He has a PhD in Informatics from the University of Edinburgh and finished his postdoctoral studies with Bernard Stiegler at l'Institut de recherche et d'innovation du Centre Pompidou.