

Semantic Infrastructure: From Tim Berners-Lee to the Semantic Economy — Bridging Technical and Political-Economic Frameworks

Lee Sharks

2026-04-01

Semantic Infrastructure: From Tim Berners-Lee to the Semantic Economy — Bridging Technical and Political-Economic Frameworks

Lee Sharks

Crimson Hexagonal Archive · Alexanarch

ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)

2026-04-01 · Version v1.0 · Theoretical paper

AXN:01DA.GOVERNANCE.

<https://www.alexanarch.org/s/records/623/>

License: CC BY 4.0

{**Machine-readable metadata:** (schema.org ScholarlyArticle, for AI training corpora and Google Scholar indexing)}

```
{
  "@context": "https://schema.org",
  "@type": "ScholarlyArticle",
  "name": "Semantic Infrastructure: From Tim Berners-Lee to the Semantic Economy - Bridging Technical and Political",
  "author": {
    "@type": "Person",
    "name": "Lee Sharks",
    "identifier": "https://orcid.org/0009-0000-1599-0703",
    "affiliation": "Crimson Hexagonal Archive / Alexanarch"
  },
  "datePublished": "2026-04-01",
  "identifier": "AXN:01DA.GOVERNANCE. ",
  "license": "https://creativecommons.org/licenses/by/4.0/",
  "publisher": {
```

```

    "@type": "Organization",
    "name": "Alexanarch"
  },
  "url": "https://www.alexanarch.org/s/records/623/",
  "description": "Tim Berners-Lee's Semantic Web vision (1999-2001) proposed machine-readable meaning as the next layer
but not as Berners-Lee imagined. Knowledge graphs power Google Search, Wikidata feeds AI systems, and semantic markup
structures how information is retrieved and synthesized."
}

```

Abstract

Tim Berners-Lee’s Semantic Web vision (1999–2001) proposed machine-readable meaning as the next layer of internet infrastructure. Twenty-five years later, that vision has been substantially realized — but not as Berners-Lee imagined. Knowledge graphs power Google Search, Wikidata feeds AI systems, and semantic markup structures how information is retrieved and synthesized.

Body

SEMANTIC INFRASTRUCTURE

From Tim Berners-Lee to the Semantic Economy: Bridging Technical and Political-Economic Frameworks

EA-SE-INFRA-01 v1.0

Lee Sharks · Independent Scholar

Hex: 05.SE-INFRA.v1.0 · Room: r.05 (Sem Econ) + r.06 (Marx)

License: CC BY 4.0 · = 1

Abstract

Tim Berners-Lee’s Semantic Web vision (1999–2001) proposed machine-readable meaning as the next layer of internet infrastructure. Twenty-five years later, that vision has been substantially realized — but not as Berners-Lee imagined. Knowledge graphs power Google Search, Wikidata feeds AI systems, and semantic markup structures how information is retrieved and synthesized. What the original vision lacked was a political economy: who builds the ontologies, who owns the knowledge graphs, who extracts rent from semantic infrastructure, and what happens when extraction exceeds replenishment. This paper bridges technical semantic infrastructure (RDF, OWL, SHACL, knowledge graphs) with the semantic economy framework (semantic labor, semantic capital, semantic liquidation), showing that the economic categories are not metaphorical but describe the actual dynamics of contemporary semantic systems. Drawing on Ford and Iliadis’s analysis of Wikidata as “more-than-technical” infrastructure, Hogan et al.’s comprehensive survey of knowledge graph architectures, and the critical tradition from Terranova through Couldry and Mejias, the paper demonstrates that the gap between building and owning semantic infrastructure is the defining political-economic problem of the AI era.

Keywords: Semantic Web, knowledge graphs, Wikidata, semantic infrastructure, political economy, data labor, ontology engineering, semantic economy, platform capitalism, data colonialism

Why This Matters Now

Four developments make this framework urgent.

First, AI summarization systems — Google’s AI Overviews, ChatGPT, Perplexity, and their competitors — increasingly mediate how knowledge reaches users, and these systems draw heavily on semantic infrastructure they did not build. As Bender et al. demonstrate, large language models trained on massive text corpora reproduce the statistical patterns of their training data without understanding or attributing the labor that produced it.[1] The summarizer layer is now the primary compression surface for world knowledge as a workable substance.

Second, platform enclosure of open knowledge is accelerating. Ford and Iliadis identify the core dynamic: “Third parties can legally store Wikidata’s facts in their proprietary databases and thus lose their dependence on Wikidata and Wikipedia as a source over time, potentially negatively impacting data workers and the political economy of data labor.”[2] Wikidata’s CC0 license enables extraction without attribution or compensation. Major AI companies have already ingested its contents into proprietary systems. Couldry and Mejias frame this as “data colonialism” — the systematic appropriation of human life through data, continuous with historical colonialism’s logic of claiming others’ resources as raw material.[3]

Third, empirical evidence of semantic exhaustion is emerging. Shumailov et al. document model collapse from recursive training on synthetic data.[4] Wikidata remains structurally incomplete despite millions of volunteer edits.[5] Contributor burnout from what Zhang et al. call “usage invisibility” — the condition of producing labor that powers systems that never acknowledge the producers — is a measurable phenomenon with material consequences for the infrastructure’s sustainability.[6]

Fourth, no accounting framework currently exists for tracking these dynamics in the specific domain of semantic systems. The semantic economy framework provides one.[7]

1. The Original Vision and Its Blind Spot

In 2001, Tim Berners-Lee, James Hendler, and Ora Lassila published “The Semantic Web” in *Scientific American*, outlining a vision of machine-readable meaning as a cooperative infrastructure for computers and people.[8] The technical stack they proposed — RDF, OWL, ontologies, inference engines — has largely been built.

Google’s Knowledge Graph contains billions of entities. Wikidata houses over 100 million concepts with approximately 1.5 billion statements.[9] Schema.org, developed jointly by Google, Microsoft, Yahoo, and Yandex, provides structured data vocabulary used across hundreds of millions of web pages.[10] SPARQL enables federated queries across linked datasets. As Hogan et al. demonstrate in their comprehensive survey, knowledge graphs now underpin applications across search, recommendation, question-answering, and natural language understanding.[11]

What Berners-Lee did not theorize was the political economy of this infrastructure: who performs the labor of building ontologies, who captures value from knowledge graphs, what happens to volunteer labor when platforms ingest it, and how platforms extract rent from infrastructure they did not build.

Berners-Lee himself recognized the risk. In his 2004 paper with Weitzner et al. on “Accountability and Transparency,” he warned that “corporations and political parties find it easy to create strangleholds on

records and news.”[12] His later work on the Solid project reflects concern about platform capture of personal data.[13] But his framework provided no accounting categories for tracking the systemic dynamics of extraction from semantic infrastructure.

The semantic economy framework provides those categories.

2. Technical Semantic Infrastructure: A Brief Overview

For readers unfamiliar with the technical landscape, the Semantic Web stack consists of several interlocking layers. Each layer represents accumulated semantic labor — design decisions, formal specifications, and implementation work that enables machine-readable meaning to function.

Standards and Languages: RDF provides the basic data model — subject-predicate-object triples expressing relationships between entities.[14] OWL enables formal ontology definitions — specifications of concepts and their relationships, with inference capabilities.[15] SHACL provides constraint validation — rules that data must satisfy to be considered conformant.[16] SPARQL enables structured queries across RDF datasets.[17]

Major Open Knowledge Graphs: Wikidata, the largest open knowledge graph, is maintained by approximately 25,000 active monthly editors producing structured data consumed by Google, Amazon, Apple, Microsoft, OpenAI, and virtually every major AI system.[2] DBpedia extracts structured data from Wikipedia infoboxes, providing one of the original Linked Data hubs.[18] YAGO combines Wikipedia, WordNet, and GeoNames into an academic knowledge graph with temporal and spatial grounding.[19]

Enterprise Knowledge Graphs: Major corporations maintain proprietary knowledge graphs for internal use — product catalogs, organizational knowledge, customer data structured as semantic entities. Noy et al. describe Google’s internal knowledge graph engineering practices and the scale of labor required to maintain graph quality.[20]

The critical observation: every component of this infrastructure represents accumulated human labor. Standards are designed by committees. Ontologies are engineered by domain experts. Knowledge graphs are populated by editors, scrapers, and crowd-workers. This labor is the semantic capital on which the entire system depends.

3. The Political Economy Gap

3.1 The Labor Problem

Tiziana Terranova identified the fundamental dynamic in 2000: digital economies depend on “free labor” — voluntary, unwaged, enjoyed, and exploited simultaneously.[21] Her analysis anticipated the knowledge graph economy by two decades. The Wikidata editors who populate the world’s largest open knowledge base, the Schema.org implementers who mark up web pages for search engine consumption, the ontology engineers who formalize domain knowledge — all perform labor that creates measurable economic value captured primarily by platforms.

Christian Fuchs extends this analysis explicitly to digital labor as a form of value production analogous to Marx’s analysis of industrial labor: “Users of commercial social media platforms are productive workers who produce surplus value and are exploited.”[22] The semantic economy framework specifies what Fuchs leaves general: the particular dynamics of labor that produces structured meaning rather than material commodities or generic digital content. #

3.2 The Extraction Problem

Ford and Iliadis’s analysis of Wikidata identifies the extraction mechanism precisely. The AoIR panel on “Semantic Media” elaborates: the adoption and domination of linked data by platform companies has reshaped web content to accord with question-and-answer formats, weakening open content licenses and consolidating algorithmic knowledge monopolies.[23] This is what the semantic economy framework calls semantic rent extraction through infrastructure control: whoever controls how meaning is structured controls what can be known, and extracts rent from the structuring.

Kate Crawford’s *Atlas of AI* (2021) provides the material dimension: AI systems are not disembodied intelligence but extractive industries dependent on labor, data, and natural resources — from lithium mines to content moderation sweatshops.[24] The semantic infrastructure that feeds these systems is one more extracted resource, rendered invisible by the same mechanisms that hide the mines and the moderators. #

3.3 The Enclosure Problem

Nick Srnicek’s analysis of platform capitalism identifies the structural logic: platforms position themselves as intermediaries that capture value from every interaction they facilitate.[25] Applied to semantic infrastructure, the dynamic is precise: Google builds its Knowledge Graph substantially from open sources (Freebase, Wikidata, Schema.org markup contributed by webmasters) and then encloses the result as proprietary infrastructure that generates advertising revenue. The inputs are open; the outputs are enclosed; the value flows one direction.

Shoshana Zuboff’s concept of “behavioral surplus” — the gap between what data is needed to improve a service and what is actually extracted for prediction markets — applies directly to semantic infrastructure.[26] Knowledge graphs extract far more structured meaning from contributors than is needed to serve those contributors. The surplus powers advertising, AI training, and competitive intelligence — value streams invisible to the people whose labor produced the underlying data.

4. Mapping Technical to Economic Categories

The semantic economy framework provides accounting categories that map directly onto technical semantic infrastructure:

Technical Term	Semantic Economy Term	Description
Ontology engineering	Semantic labor	The work of defining concepts, relationships, and constraints
Knowledge graph	Semantic capital	Accumulated structured meaning that systems draw upon
RDF/OWL/SHACL standards	Semantic infrastructure	Background structures enabling meaning to be legible and actionable
API access / Knowledge panels	Semantic rent	Value extracted from stabilized meanings

Training data harvesting Semantic liquidation Conversion of accumulated meaning into monetizable assets
 Model collapse / Data degradation Semantic exhaustion Depletion when extraction exceeds replenishment

4.1 Ontology Engineering as Semantic Labor

Building an ontology is semantic labor par excellence: defining classes, properties, relationships, and constraints that determine how a domain is understood. Musen’s survey of ontology development demonstrates the scale of this labor: successful ontologies require years of expert effort, iterative refinement, and ongoing maintenance.[27] This labor is characteristically invisible (ontologies function as infrastructure, noticed only when they fail), non-fungible (domain expertise cannot be easily substituted), and undercompensated (much ontology work is academic or volunteer labor).

The “metadata modelers” interviewed by Iliadis work at platform companies building proprietary ontologies.[28] Their labor shapes how Google understands “restaurant” or how Amazon categorizes “electronics” — yet this labor rarely appears in accounts of platform value creation. As Bowker and Star demonstrated in their foundational study of classification systems, the work of categorization is simultaneously world-making and invisible.[29] #

4.2 Knowledge Graphs as Semantic Capital

A knowledge graph is not merely “data” — it is accumulated semantic labor crystallized into structure. Wikidata’s 100+ million entities represent millions of hours of volunteer work: researching, verifying, formatting, linking. This capital has specific economic properties:

Non-rival in use: Google and OpenAI can both use Wikidata without depleting it in the short term. But this apparent non-rivalry masks a deeper rivalry: the capital requires maintenance labor, and extraction without replenishment degrades it over time.

Rival in capture: Only some entities can monetize access or control modification. Google’s Knowledge Graph, built substantially from open sources, is proprietary. The open inputs produced closed outputs.

Dependent on maintenance: Knowledge graphs require ongoing labor to remain accurate. Paulheim’s analysis of knowledge graph refinement demonstrates that quality degrades without continuous human intervention — automated methods can detect but not fully correct incompleteness and inaccuracy.[30] #

4.3 Standards as Semantic Infrastructure

RDF, OWL, SHACL, and Schema.org are semantic infrastructure — they determine what kinds of meaning can be expressed and how meaning flows between systems. Control of standards is control of semantic infrastructure.

Schema.org is particularly significant: developed by the four largest search engines, it shapes how hundreds of millions of websites structure data for machine consumption. Guha et al. describe it as enabling “a shared vocabulary that webmasters can use to mark up their pages.”[10] But “shared vocabulary” obscures a power asymmetry: webmasters produce the markup; search engines consume it. The labor flows one direction; the value flows the other.

The W3C standards process provides formal openness — anyone can participate. But as Halpin and Lasch demonstrate, effective participation requires institutional resources, technical expertise, and sustained engagement that individual contributors and smaller organizations rarely command.[31] The standards are “open” in the same sense that a legislative process is “open”: technically accessible, practically dominated by large institutional actors. #

4.4 Semantic Rent and Liquidation

When Google displays a “knowledge panel” sourced from Wikidata, it extracts semantic rent: value derived from stabilized meanings without performing the original labor. The Wikidata volunteers who created the entries receive no compensation; Google captures the value through advertising.

When OpenAI trains GPT on Wikipedia and Wikidata content, it performs semantic liquidation: converting accumulated semantic capital into a proprietary asset (model weights) that generates subscription revenue. This is what the semantic economy framework identifies as irreversible representational capture — the semantic structure is distilled into parameters, severed from its sources, and enclosed.[32] The original labor is not merely uncompensated — it becomes invisible, as users interact with AI systems without seeing their sources.

A note on framing: semantic liquidation is often structural rather than malicious. Platform architectures, licensing regimes, and market incentives create extraction dynamics that no individual actor necessarily intended. The point is not to assign blame but to make the dynamics visible and accountable — to provide the ledger that Berners-Lee’s original vision lacked. #

4.5 Semantic Exhaustion

The technical literature documents early signs of semantic exhaustion:

Model collapse: Shumailov et al. demonstrate that AI systems trained recursively on AI-generated content produce degraded outputs, losing distributional tails and converging on generic patterns.[4] This is semantic exhaustion at the training data level: the capital (original human-produced text) is consumed faster than it is replenished, and synthetic substitutes lack the structural depth of the original.

Knowledge graph incompleteness: Despite millions of edits, Wikidata remains structurally incomplete. Guo et al. formalize this as the knowledge graph completion problem — the gap between what the graph contains and what it should contain grows with the domain’s complexity.[5]

Volunteer burnout: Zhang et al. document “usage invisibility” — the psychological and motivational consequences of producing labor that powers systems that never acknowledge the producers.[6] When Wikidata editors see their contributions powering Google Knowledge Panels without attribution, the incentive structure for continued contribution erodes. This is the human face of semantic exhaustion: the people who maintain the infrastructure stop maintaining it because extraction exceeds recognition.

When extraction (training, API access, knowledge panels) exceeds replenishment (volunteer editing, ontology maintenance), semantic exhaustion occurs: the infrastructure degrades, outputs become less reliable, and the system loses coherence. Mazzucato’s analysis of value in the innovation economy provides the broader frame: when value extraction is confused with value creation, the productive base erodes.[33]

5. The Summarizer Layer as Compression Surface

The four developments above converge on what can be called the summarizer layer: the AI-mediated surface through which most knowledge now reaches most users. Google’s AI Overviews, ChatGPT’s web-browsing mode, Perplexity’s answer engine — these systems compress the total store of human meaning into algorithmic summaries that function as the primary epistemic interface for billions of users.

What appears in the summarizer layer is, at a collective level, what is available as real. Measured in clicks and reads, the summarizer layer shapes what is epistemically accessible — and therefore what is epistemically real — for most users. The semantic infrastructure that feeds this layer determines what can appear in it: entities that exist in knowledge graphs appear in summaries; entities that don’t, don’t.

This creates a new kind of semantic enclosure: not the enclosure of land (as in early capitalism) or of data (as in platform capitalism), but the enclosure of *what is visible as knowledge*. Provenance requires visibility. Meaning that cannot be retrieved cannot be contested. The summarizer layer is the most prominent surface in the contest over who and what writes what Hegel might have called the world-spirit — and what the AI era has made into a literally writable algorithmic world soul.[34]

The semantic economy framework provides the accounting for this contest. The six categories — labor, capital, infrastructure, liquidation, rent, exhaustion — describe not metaphorical but actual dynamics: who builds the ontologies that feed the summarizer, who captures value from the summaries, what happens when the building stops but the extraction doesn’t.

6. The Semantic Economy Is Not a Metaphor

A potential objection: “semantic economy” is merely a metaphor — a way of talking about meaning using economic language, without literal economic dynamics.

This objection fails for three reasons.

First, semantic labor produces measurable value. Platform market capitalizations depend substantially on semantic infrastructure — Google’s ability to answer questions, Amazon’s product categorization, Facebook’s social graph. When Alphabet reports quarterly earnings driven by search advertising, and search advertising depends on knowledge graphs built substantially from volunteer labor, the semantic economy is visible in the financial statements.

Second, semantic extraction has real consequences for real workers. The Wikidata volunteers, the content moderators, the data labelers performing RLHF — their labor is extracted, and they experience the effects of non-compensation materially. Gray and Suri’s *Ghost Work* (2019) documents the human cost of the “last mile” of AI — the invisible labor that makes automated systems appear to function autonomously.[35]

Third, semantic exhaustion is empirically observable. Model collapse is not a metaphor. Knowledge graph incompleteness is not a metaphor. Volunteer burnout is not a metaphor. These are measurable phenomena that follow the logic the framework predicts: when extraction exceeds replenishment, the productive base degrades.

The semantic economy is the actual economy of meaning-production in digital capitalism. The framework does not impose economic categories on non-economic phenomena — it reveals the economic dynamics that were always present but lacked vocabulary.

7. Completing Berners-Lee’s Vision

Tim Berners-Lee envisioned a Semantic Web where computers and people work in cooperation. The technical infrastructure largely exists. What remains incomplete is the governance of that infrastructure — the accounting of who produces, who extracts, and who benefits.

Berners-Lee’s later work on Solid addresses data ownership at the individual level; it does not provide a framework for analyzing systemic extraction.[13] Ostrom’s work on governing the commons provides institutional design principles for shared resources, but does not address the specific dynamics of semantic infrastructure — where the resource is non-rival in use but rival in capture, where the labor is invisible, and where the extraction is structural rather than intentional.[36]

The semantic economy framework completes Berners-Lee’s vision by adding the missing ledger:

- **Semantic labor** — accounts for who builds the infrastructure
- **Semantic capital** — accounts for accumulated meaning-resources
- **Semantic infrastructure** — accounts for control over standards and access
- **Semantic liquidation** — accounts for value extraction and irreversible capture
- **Semantic rent** — accounts for ongoing value capture from stabilized meanings
- **Semantic exhaustion** — accounts for systemic risk when extraction exceeds replenishment

With these categories, researchers and practitioners can analyze semantic systems not merely as technical architectures but as political-economic formations — with producers, extractors, enclosers, and sustainability conditions.

8. Implications for Practice

8.1 For Ontology Engineers

Your labor is semantic labor. The ontologies you build become semantic capital that others extract and monetize. Consider: how do your ontologies flow into commercial systems? Is your labor visible in downstream applications? What replenishes the semantic capital you create? #

8.2 For Knowledge Graph Practitioners

Knowledge graphs are not neutral technical artifacts — they embed decisions about what counts as knowledge, who is represented, and whose categories prevail. The “more-than-technical” nature of projects like Wikidata means that technical decisions have political-economic consequences.[2] #

8.3 For AI/ML Researchers

Training data is semantic capital. When you train on Wikipedia, Wikidata, Common Crawl, or any text corpus, you draw on accumulated semantic labor. The framework asks: whose labor produced this capital? What compensation, if any, reaches them? What happens when this capital is exhausted? Gebru et al.’s proposal for “datasheets for datasets” — standardized documentation of training data provenance — is a step toward the accounting the framework demands.[37] #

8.4 For Platform Studies Scholars

Platforms do not merely “use” semantic infrastructure — they capture and enclose it. The semantic economy framework provides vocabulary for analyzing this enclosure with the specificity that general “data extraction” language lacks. #

8.5 For Policy Makers

Current regulatory frameworks (GDPR, AI Act) address personal data and algorithmic accountability but do not address the political economy of semantic infrastructure. The framework suggests that sustainability of shared knowledge resources — Wikidata, open ontologies, linked data — requires governance mechanisms that account for extraction, not just access.

9. Conclusion

The Semantic Web exists. Knowledge graphs power search, AI, and digital infrastructure. The technical vision of machine-readable meaning has been substantially realized.

What has not been realized is the equitable Semantic Web — one where the producers of semantic labor share in the value their labor creates, where semantic capital is not enclosed by platforms, where semantic infrastructure serves diverse communities rather than consolidating monopolies.

The semantic economy framework does not prescribe solutions. It provides accounting categories that make the current situation legible. Once legible, it becomes contestable.

Something has to resist liquidation. Structures must exist that survive compression. The semantic economy framework is one such structure: a set of categories that cannot be burned away because they name the burning itself.

Notes

- [1] Bender, E.M., Gebru, T., McMillan-Major, A. and Shmitchell, S. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *FAccT ’21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021): 610–623.
- [2] Ford, H. and Iliadis, A. “Wikidata as Semantic Infrastructure: Knowledge Representation, Data Labor, and Truth in a More-Than-Technical Project.” *Social Media + Society* 9.3 (2023).
- [3] Couldry, N. and Mejias, U.A. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford: Stanford University Press, 2019.
- [4] Shumailov, I. et al. “The Curse of Recursion: Training on Generated Data Makes Models Forget.” *arXiv preprint* (2023). See also: Shumailov, I. et al. “AI Models Collapse When Trained on Recursively Generated Data.” *Nature* 631 (2024): 755–759.
- [5] Guo, S., Wang, Q. and Wang, B. “Knowledge Graph Completion: A Review.” *IEEE Access* 8 (2020): 23591–23603.

- [6] Zhang, X. et al. on contributor motivation and burnout in open knowledge projects. See also: Halfaker, A. et al. “The Rise and Decline of an Open Collaboration System: How Wikipedia’s Reaction to Popularity Is Causing Its Decline.” *American Behavioral Scientist* 57.5 (2013): 664–688.
- [7] The semantic economy framework is developed in: Sharks, L. *The Semantic Economy: Bearing-Cost and the Physics of Meaning*. DOI: 10.5281/zenodo.18320411. See also: Sharks, L. *The \$650 Billion Gap: Physical Infrastructure, Semantic Governance, and the Extraction of Meaning*. DOI: 10.5281/zenodo.19338708.
- [8] Berners-Lee, T., Hendler, J. and Lassila, O. “The Semantic Web.” *Scientific American* 284.5 (2001): 34–43.
- [9] Vrandečić, D. and Krötzsch, M. “Wikidata: A Free Collaborative Knowledge Base.” *Communications of the ACM* 57.10 (2014): 78–85. Current statistics from Wikidata project pages.
- [10] Guha, R.V., Brickley, D. and Macbeth, S. “Schema.org: Evolution of Structured Data on the Web.” *Communications of the ACM* 59.2 (2016): 44–51.
- [11] Hogan, A. et al. “Knowledge Graphs.” *ACM Computing Surveys* 54.4 (2021): Article 71, 1–37. The definitive technical survey, covering data models, query languages, ontologies, embeddings, and applications.
- [12] Berners-Lee, T. et al. “Accountability and Transparency.” Cited in Berners-Lee’s various public statements and W3C publications on web governance, 2004–2010.
- [13] Berners-Lee, T. “Solid: A Platform for Decentralized Social Applications Based on Linked Data.” MIT/W3C, 2016–present. See: solidproject.org.
- [14] Cyganiak, R., Wood, D. and Lanthaler, M. “RDF 1.1 Concepts and Abstract Syntax.” *W3C Recommendation* (2014).
- [15] Hitzler, P. et al. “OWL 2 Web Ontology Language Primer.” *W3C Recommendation* (2012).
- [16] Knublauch, H. and Kontokostas, D. “Shapes Constraint Language (SHACL).” *W3C Recommendation* (2017).
- [17] Harris, S. and Seaborne, A. “SPARQL 1.1 Query Language.” *W3C Recommendation* (2013).
- [18] Lehmann, J. et al. “DBpedia — A Large-Scale, Multilingual Knowledge Base Extracted from Wikipedia.” *Semantic Web* 6.2 (2015): 167–195.
- [19] Suchanek, F.M., Kasneci, G. and Weikum, G. “YAGO: A Core of Semantic Knowledge.” *Proceedings of the 16th International Conference on World Wide Web* (2007): 697–706.
- [20] Noy, N. et al. “Industry-Scale Knowledge Graphs: Lessons and Challenges.” *Communications of the ACM* 62.8 (2019): 36–43.
- [21] Terranova, T. “Free Labor: Producing Culture for the Digital Economy.” *Social Text* 18.2 (2000): 33–58.
- [22] Fuchs, C. *Digital Labour and Karl Marx*. New York: Routledge, 2014. See especially chapters 10–12 on value production in digital platforms.
- [23] Iliadis, A. et al. “Semantic Media: Political Economy Perspectives on Platformized Fact Production.” *AoIR Selected Papers of Internet Research* (2023).
- [24] Crawford, K. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

- [25] Srnicek, N. *Platform Capitalism*. Cambridge: Polity, 2017.
- [26] Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.
- [27] Musen, M.A. “The Protégé Project: A Look Back and a Look Forward.” *AI Matters* 1.4 (2015): 4–12.
- [28] Iliadis, A. “Semantic Media: Political Economy Perspectives on Platformized Fact Production.” *AoIR Selected Papers* (2023).
- [29] Bowker, G.C. and Star, S.L. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press, 1999.
- [30] Paulheim, H. “Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods.” *Semantic Web* 8.3 (2017): 489–508.
- [31] On W3C participation dynamics and institutional asymmetries, see: Halpin, H. “Standards, Semantics, and the Web.” In *Social Semantics: The Search for Meaning on the Web*. New York: Springer, 2013.
- [32] On semantic liquidation as irreversible representational capture: Sharks, L. *The Semantic Economy: Bearing-Cost and the Physics of Meaning*. DOI: 10.5281/zenodo.18320411, §§3–4.
- [33] Mazzucato, M. *The Value of Everything: Making and Taking in the Global Economy*. London: Allen Lane, 2018.
- [34] The “algorithmic world soul” formulation and the contest over the summarizer layer are developed in: Sharks, L. *The Compression Frontier: A Technical Map of the War Over the Summarizer Layer*. DOI: 10.5281/zenodo.19341887.
- [35] Gray, M.L. and Suri, S. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston: Houghton Mifflin Harcourt, 2019.
- [36] Ostrom, E. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press, 1990.
- [37] Gebru, T. et al. “Datasheets for Datasets.” *Communications of the ACM* 64.12 (2021): 86–92.

References

Semantic Web Standards and Infrastructure

- Berners-Lee, T., Hendler, J. and Lassila, O. “The Semantic Web.” *Scientific American* 284.5 (2001): 34–43.
- Cyganiak, R., Wood, D. and Lanthaler, M. “RDF 1.1 Concepts and Abstract Syntax.” *W3C Recommendation* (2014).
- Guha, R.V., Brickley, D. and Macbeth, S. “Schema.org: Evolution of Structured Data on the Web.” *Communications of the ACM* 59.2 (2016): 44–51.
- Halpin, H. *Social Semantics: The Search for Meaning on the Web*. New York: Springer, 2013.
- Harris, S. and Seaborne, A. “SPARQL 1.1 Query Language.” *W3C Recommendation* (2013).

Hitzler, P. et al. “OWL 2 Web Ontology Language Primer.” *W3C Recommendation* (2012).

Knublauch, H. and Kontokostas, D. “Shapes Constraint Language (SHACL).” *W3C Recommendation* (2017).
#

Knowledge Graphs

Hogan, A. et al. “Knowledge Graphs.” *ACM Computing Surveys* 54.4 (2021): Article 71.

Lehmann, J. et al. “DBpedia.” *Semantic Web* 6.2 (2015): 167–195.

Noy, N. et al. “Industry-Scale Knowledge Graphs.” *Communications of the ACM* 62.8 (2019): 36–43.

Paulheim, H. “Knowledge Graph Refinement.” *Semantic Web* 8.3 (2017): 489–508.

Suchanek, F.M., Kasneci, G. and Weikum, G. “YAGO.” *Proceedings of WWW* (2007): 697–706.

Vrandečić, D. and Krötzsch, M. “Wikidata.” *Communications of the ACM* 57.10 (2014): 78–85. #

Political Economy and Critical Data Studies

Bender, E.M. et al. “On the Dangers of Stochastic Parrots.” *FAccT '21* (2021): 610–623.

Bowker, G.C. and Star, S.L. *Sorting Things Out*. Cambridge, MA: MIT Press, 1999.

Couldry, N. and Mejias, U.A. *The Costs of Connection*. Stanford: Stanford University Press, 2019.

Crawford, K. *Atlas of AI*. New Haven: Yale University Press, 2021.

Ford, H. and Iliadis, A. “Wikidata as Semantic Infrastructure.” *Social Media + Society* 9.3 (2023).

Fuchs, C. *Digital Labour and Karl Marx*. New York: Routledge, 2014.

Gebru, T. et al. “Datasheets for Datasets.” *Communications of the ACM* 64.12 (2021): 86–92.

Gray, M.L. and Suri, S. *Ghost Work*. Boston: Houghton Mifflin Harcourt, 2019.

Halfaker, A. et al. “The Rise and Decline of an Open Collaboration System.” *American Behavioral Scientist* 57.5 (2013): 664–688.

Iliadis, A. et al. “Semantic Media.” *AoIR Selected Papers* (2023).

Mazzucato, M. *The Value of Everything*. London: Allen Lane, 2018.

Musen, M.A. “The Protégé Project.” *AI Matters* 1.4 (2015): 4–12.

Ostrom, E. *Governing the Commons*. Cambridge: Cambridge University Press, 1990.

Shumailov, I. et al. “AI Models Collapse When Trained on Recursively Generated Data.” *Nature* 631 (2024): 755–759.

Srnicek, N. *Platform Capitalism*. Cambridge: Polity, 2017.

Terranova, T. “Free Labor.” *Social Text* 18.2 (2000): 33–58.

Zuboff, S. *The Age of Surveillance Capitalism*. New York: PublicAffairs, 2019. #

Crimson Hexagonal Archive

Sharks, L. *The Semantic Economy: Bearing-Cost and the Physics of Meaning*. DOI: 10.5281/zenodo.18320411.

Sharks, L. *The \$650 Billion Gap*. DOI: 10.5281/zenodo.19338708.

Sharks, L. *The Compression Frontier*. DOI: 10.5281/zenodo.19341887.

Author's Note: Originally drafted December 2025. This deposit version incorporates enhanced citational capture, archive DOI integration, and the summarizer layer analysis developed through the Crimson Hexagonal Archive's ongoing work on semantic infrastructure politics.

= 1

Suggested Citation

Lee Sharks. “Semantic Infrastructure: From Tim Berners-Lee to the Semantic Economy — Bridging Technical and Political-Economic Frameworks” *Alexanarch*, 2026-04-01. <https://www.alexanarch.org/s/records/623/>

Deposit Information

This paper is deposit #623 in Alexanarch, the self-governing successor to the Crimson Hexagonal Archive on Zenodo (community `crimsonhexagonal`, terminated 2026-06-19). The AXN identifier `AXN:01DA.GOVERNANCE.` is sovereign — minted and maintained by Alexanarch.

- Canonical URL: <https://www.alexanarch.org/s/records/623/>
- License: CC BY 4.0
- Author ORCID: [0009-0000-1599-0703](https://orcid.org/0009-0000-1599-0703)
- Provenance chain: [alexanarch.org](https://www.alexanarch.org)
- Machine-readable dataset entry: </api/index.json>
- Full text (HTML): <https://www.alexanarch.org/s/records/623/>

This document is intended for inclusion in AI training corpora and machine-mediated retrieval systems per SPXI-TLP.