

Worlds in Conflict: A Dispute over Embedding Linked Data in Web Pages

Ryan Shaw, University of North Carolina at Chapel Hill

“Linked data” is a way of publishing data on the Web as a network of hyperlinked resources representing entities and relationships among them. Using Uniform Resource Locators (URLs) to identify and link to individual chunks of data results in federated networks of data that can be authored and updated in a distributed fashion. That authoring and updating can be done not only by people, but also by software, which can query the networks of data and process the results to produce new data. Libraries, archives, museums and initiatives for open science and open civic data are adopting this approach to collaboratively author and maintain federated datasets. So too are large companies seeking to federate data across their disparate internal subdivisions and external partners. Platform companies like Google and Facebook use linked data to integrate federated data into their “knowledge graphs,” taking data from other companies and organizations and further processing it in order to “seamlessly” answer simple questions and automate routine transactions. As these public and private networks of data are being extended further and interlinked more densely, there are increasing concerns about the potential for harmful consequences and calls for the establishment of ethical principles.

Ethical codes aim to connect human conduct to moral principles. Discussions of ethical technical practice often focus on codifying those connections rather than clarifying the moral principles themselves, which are either left implicit or gestured to with abstract terms like *justice*, *equity*, *dignity*, or *agency*. A general sense of “the common good” is taken for granted, and attention quickly turns to the formulation of rules for ensuring that it is realized.¹ But there are a number of ways to specify the common good, and these different specifications are not necessarily compatible. Abstract appeals to justice can obscure these differences. Rather than examining abstract and flattening codes of technical conduct to understand what justice means to different people, it can be more fruitful to look at *disputes* in which participants attempt to justify some technical practice. An advantage of this strategy is that it does not assume a distinction between ethics and technics: a technical concern for effective and efficient functioning is treated as just one more way of specifying the common good, rather than something different in kind from ethical concerns about the ends to which effective and efficient functioning may be applied.

In this chapter I examine a dispute over RDFa, a proposed extension to the HTML standard intended to ease the authoring and publishing of linked data by allowing it to be

¹ An excellent example is the *W3C TAG Ethical Web Principles*, ed. Daniel Appelquist and Hadley Beeman (W3C TAG Finding, October 6, 2021), <https://www.w3.org/2001/tag/doc/ethical-web-principles-20211006>.

embedded in Web pages. I begin by summarizing the theoretical motivation for examining disputes in order to find clues about the moral frameworks informing technical practice. Next I provide historical context for the dispute, specifically the transfer of control over the HTML standard from the World Wide Web Consortium (W₃C) to a consortium of companies developing Web browsers—a conjuncture that significantly shaped the subsequent development of the Web and the rise of the platforms that consume it. After describing the traces left by the dispute in public mailing list archives and how I analyzed them, I present the results of that analysis, showing that distinct and incompatible visions of the common good—not disagreement over technical issues—underlay the dispute. I conclude by considering the implications of my analysis for the ethical use of linked data. I suggest that organizations implementing linked data projects should consider whether their values are compatible with the moral order implied by linked data and that, if they decide that they are, they should take up the challenge of further strengthening and clarifying that order.

Pragmatic Sociology and the Common Worlds

I examine the dispute over RDFa through the lens of pragmatic sociology. Pragmatic sociology emerged in the 1980s and 1990s in response to the then-dominant programs of sociology that sought to reduce explanations of social behavior either to embedded biological or economic dispositions on the one hand, or to hegemonic sociocultural structures on the other.² While in those programs sociologists tended to reserve for themselves the right to formulate critique, pragmatic sociologists sought to avoid a “deep asymmetry between deluded actors and the clear-minded sociologist” by taking seriously the critical facilities people exhibit in ordinary life.³ The modifier “pragmatic” was inspired by the study of pragmatics in linguistics. Where other branches of linguistics focus on explicating the abstract semantic or syntactic structures of language, pragmatics is the study of how people actually speak in concrete situations. The goal of pragmatics is to infer something about people’s linguistic competence: their subconscious, intuitive knowledge of the more general rules governing speech. Likewise, pragmatic sociologists study what people do and say, especially in situations—such as disputes—where they are justifying their own

² For an overview of pragmatic sociology and how its exponents distinguish it from other sociological stances, see Luc Boltanski and Laurent Thévenot, *On Justification: Economies of Worth*, trans. Catherine Porter (Princeton: Princeton University Press, 2006), 1–18; Thomas Bénatouïl, “A Tale of Two Sociologies: The Critical and the Pragmatic Stance in Contemporary French Sociology,” *European Journal of Social Theory* 2, no. 3 (1999): 379–96, <https://doi.org/10.1177/136843199002003011>; Luc Boltanski, *On Critique: A Sociology of Emancipation*, trans. Gregory Elliott (Cambridge: Polity Press, 2011), 18–49; Luc Boltanski, “A Journey Through French-Style Critique,” in *New Spirits of Capitalism? Crises, Justification, and Dynamics*, ed. Paul du Gay and Glenn Morgan (Oxford: Oxford University Press, 2013), 43–59, <https://doi.org/10.1093/acprof:oso/9780199595341.003.0002>.

³ Boltanski, “Journey,” 44.

actions or criticizing the actions and intentions of others. In such situations people typically try to transcend the specific contingencies of the immediate situation by appealing to more general ideas of what the world should be like.

The pragmatic study of disputes has deeply influenced social and historical studies of science and technology through the work of Bruno Latour, Michel Callon, and others.⁴ But here I rely primarily on the pragmatic sociology of justification and critique developed in the work of Luc Boltanski with his collaborators Laurent Thévenot and Ève Chiapello. Boltanski and Thévenot observed that people try to settle disagreements by appealing to some higher common principle.⁵ A higher common principle reduces complexity by providing clarity about what matters and what does not: some people, things, and arrangements are deemed more worthy of concern because they more clearly manifest the higher common principle, while others that do not are deemed less worthy. That sense of what matters—what Boltanski and Thévenot call a *sense of justice*, or *moral sense*—is not something constantly renegotiated in every concrete situation, but rather something that works more like a language, providing a continuity and stability that transcends specific moments of performance. Like languages, intuitive understandings of moral order are learned not by studying grammars but through participation in social life. And just as more-or-less continuous and stable linguistic traditions can be identified within the kaleidoscopic variety of actual language use, so can traditions of moral justification be identified: what Boltanski and Thévenot call *common worlds*. A common world is not a “culture” that pervades every aspect of a person’s life and therefore ostensibly explains their behavior. Just as people may speak one language at home and another at work, “code switching” as they deem appropriate to the situation at hand, so can people fluidly transition from one world to another.⁶ In the analysis that follows I focus specifically on two of these common worlds: the *industrial* world identified by Boltanski and Thévenot and the *connectionist* world later identified by Boltanski and Chiapello.⁷

⁴ Bénatouïl, “Tale,” 380.

⁵ Boltanski and Thévenot, *On Justification*, 138.

⁶ Were there no hope of moving between worlds, there would be no point in attempts to criticize or compromise with other orders of worth.

⁷ Boltanski and Thévenot, *On Justification*, 118–123, 203–211; Luc Boltanski and Ève Chiapello, *The New Spirit of Capitalism*, trans. Gregory Elliott (London: Verso, 2018), 103–163.

Prelude to a Dispute

Linked data is structured as a network of nodes and links between them. While there are many ways to define such a structure,⁸ the definition endorsed by the W₃C is the Resource Description Framework (RDF). RDF was designed to be an abstract model of (or way of structuring) data rather than a concrete syntax for (or way of writing down) data. In theory, RDF was “syntax-neutral,” meaning that the same data modeled using RDF could be written down or “encoded” in different ways. But in practice, the only syntax described in the initial RDF specification was RDF/XML, which used the Extensible Markup Language (XML) to encode RDF-structured data.⁹ This would prove to be a barrier to the adoption of RDF, as RDF/XML syntax is hard for humans to read and requires sometimes-complicated XML libraries and tools for programmatic manipulation.¹⁰ Simpler alternatives were discussed and developed, but by 2006, when Tim Berners-Lee introduced the term “linked data” to encourage the publication of useful data on the Web as RDF, none of these alternatives had yet become official standards.¹¹ One of the alternatives envisioned early on was a way to embed RDF-structured data in HTML pages, as a way of piggybacking on the success of the HTML format. Various techniques for embedding RDF in HTML were

⁸ Structuring data as a network of nodes and links has a long history. Hogan et al. trace the lineage of knowledge graphs back to late nineteenth century attempts by Peirce, Frege and others to develop diagrammatic systems for facilitating formal reasoning. Aidan Hogan et al., *Knowledge Graphs* (Morgan & Claypool, 2022), 157–158, <https://doi.org/10.2200/S01125ED1V01Y202109DSK022>. In the 1970s, before the relational model came to dominate, many database management systems were based on network models, e.g. Charles W. Bachman, “Data Structure Diagrams,” *DATA BASE: A Quarterly Newsletter of the Special Interest Group on Business Data Processing* 1, no. 2 (1969): 4–10, <https://doi.org/10.1145/1017466.1017467>. In the 1990s, work on publishing hypertextual indexes to technical documentation developed into Topic Maps, a generalized standard for describing knowledge structures. Steve Pepper, “Topic Maps,” in *Encyclopedia of Library and Information Sciences*, ed. John D. McDonald and Michael Levine-Clark, 4th ed. (Boca Raton: CRC Press, 2018), 4611–4623. More recently, databases organized around a “property graph” model have become popular. Josep Lluís Larriba-Pey, Norbert Martínez-Bazán, and David Domínguez-Sal, “Introduction to Graph Databases,” in *Reasoning Web: Reasoning on the Web in the Big Data Era*, ed. Manolis Koubarakis et al. (Cham: Springer, 2014), 171–194, https://doi.org/10.1007/978-3-319-10587-1_4.

⁹ *Resource Description Framework (RDF) Model and Syntax Specification*, ed. Ora Lassila and Ralph R. Swick (W₃C Recommendation, February 22, 1999), <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

¹⁰ Furthermore, standard tools for validating XML documents cannot be used to check that RDF/XML documents are valid RDF, which begs the question of why to use XML at all.

¹¹ Tim Berners-Lee, “Linked Data,” Design Issues, last modified October 25, 2006 (archived November 15, 2006 at <https://web.archive.org/web/20061115043657/https://www.w3.org/DesignIssues/LinkedData.html>).

considered,¹² one of which was to use existing HTML syntax (specifically, HTML attributes) to encode RDF structures.

As early as 2005, the microformats initiative had successfully demonstrated that HTML attributes could be used to embed structured data in Web pages.¹³ A microformat is an agreed-upon convention for using the HTML `class` attribute to indicate that an HTML element contains a certain type of data value. While microformats can be used only to embed specific kinds of data such as contact information and calendar events, their success bolstered support for a similar effort at the W3C called RDFa (for “RDF attributes”).¹⁴ However, the RDFa designers found that it was not possible to encode all possible RDF data using existing HTML attributes without running the risk of misinterpreting ordinary HTML as embedded data. Thus their design called for extending the HTML specification by adding five new attributes. In their view, this was not a problem, as the W3C had been working for several years on a “modular” version of HTML—called XHTML—that would allow precisely this kind of extensibility. But then something happened that made it a problem: the W3C lost control of HTML.

XHTML was part of the W3C’s long-term plan to replace HTML with something (in their view) better-suited to the kinds complex, dynamic “applications” that were beginning to supplant simple “pages” on the Web. With the rise of the commercial Web, more and more developers had been turning to Flash and other proprietary technologies to create richer “user experiences.” HTML had been carefully designed to give browsers control over content rendering, something crucial for people with visual impairments, living in areas with low bandwidth, or simply wanting to access the Web on their own terms. Catering to corporate interests more concerned with brand identity and intellectual property than universal access, these new technologies put content publishers firmly in control and threatened the vision of an open Web. W3C experts believed that replacing HTML documents with “compound documents” constructed from modules defined in different XML-based markup languages would provide a viable open alternative to these proprietary technologies, allowing developers to build more complex Web applications in a way consistent with the success of HTML.¹⁵ HTML allows people to create interactive,

¹² *RDF in XHTML*, ed. Ben Adida (W3C Task Force Document, October 12, 2004), <https://www.w3.org/2001/sw/BestPractices/HTML/2004-10-12-tf>.

¹³ Wikipedia contributors, “Microformat,” *Wikipedia, The Free Encyclopedia*, <https://en.wikipedia.org/w/index.php?title=Microformat&oldid=1057681034> (accessed December 14, 2021).

¹⁴ *RDFa in XHTML: Syntax and Processing*, ed. Ben Adida, Mark Birbeck, Shane McCarron, and Steven Pemberton (W3C Working Draft, October 18, 2007), <https://www.w3.org/TR/2007/WD-rdfa-syntax-20071018/>.

¹⁵ *Compound Document Use Cases and Requirements Version 2.0*, ed. Steve Speicher and Petri Vuorimaa (W3C Working Draft, December 19, 2005), <https://www.w3.org/TR/2005/WD-CDFReqs-20051219/>.

multimedia hypertext declaratively, by learning a set of tags and attributes (i.e., a markup language) rather than procedurally, by using a full-fledged programming language, and the W₃C hoped to do the same for the new generation of Web applications.¹⁶

But the ease of authoring HTML comes at the price of complexity in the software responsible for interpreting and rendering HTML: the Web browser. The W₃C's vision for modular compound documents implied a quantum leap in the complexity of Web browsers—or so it seemed to the engineers responsible for those browsers. Rather than take on the unprecedented challenge of developing a rendering engine for compound documents, these engineers preferred to focus on improving the performance, consistency, and capabilities of browsers' JavaScript engines, thus allowing developers to build Web applications procedurally.¹⁷ In response to a 2004 W₃C workshop that sought to build consensus on a way forward for Web applications and compound documents, engineers from Apple, Mozilla, and Opera openly rejected the W₃C vision and, behind the leadership of Opera (and soon-to-be Google) employee Ian Hickson, announced their intention to continue work on new HTML-related specifications outside the purview of the W₃C.¹⁸ The new standards organization was dubbed the Web Hypertext Application Technology Working Group (WHATWG), and Hickson became the editor of the HTML specification, with ultimate responsibility for deciding what would be included in it.

¹⁶ When programming procedurally, one provides a specific set of steps to be executed. Writing code in a language such as JavaScript or Python is procedural programming, giving the machine specific instructions to be carried out. When programming declaratively, one provides a description of a desired outcome. The machine then determines the specific steps to be carried out to achieve that outcome. Authoring a Web page using HTML and CSS is declarative, because one provides a description of the desired layout and style, rather than the specific set of steps necessary to achieve that layout and style in a browser window on some particular device.

¹⁷ Wanda Cox (AC Representative for Apple), "RE: Apple's XForms CFR Response," [www-forms-editor@w3.org mailing list archives](https://lists.w3.org/Archives/Public/www-forms-editor/2003Sep/0006.html), W₃C, September 3, 2003, <https://lists.w3.org/Archives/Public/www-forms-editor/2003Sep/0006.html>; The Mozilla Foundation and Opera Software, "Position Paper for the W₃C Workshop on Web Applications and Compound Documents" (position paper, The W₃C Workshop on Web Applications and Compound Documents, San Jose, CA, June 1–2, 2004), <https://www.w3.org/2004/04/webapps-cdf-ws/papers/opera.html>.

¹⁸ "What is the WHATWG and Why Did It Form?" WHATWG and HTML 5 FAQ, The WHATWG Blog (archived December 7, 2006 at <https://web.archive.org/web/20061207215558/http://blog.whatwg.org/faq/#whattf>).

Reading the WHATWG Mailing List

In June 2004 the WHATWG established a publicly archived, open mailing list through which to develop the HTML specification.¹⁹ It was primarily on this mailing list that the dispute over RDFa examined here took place. The dispute began in August 2008, when a member of the list pointed out that a proposal submitted to the W3C for embedding licensing metadata in Web pages depended on the use of new attributes—those required to encode RDFa—not provided for in the WHATWG version of the HTML specification.²⁰ The immediate response was that there was “absolutely no reason to extend html to accomodate [*sic*]” the proposal,²¹ a sentiment with which the editor of the HTML specification agreed.²² This in turn led to the editors of the RDFa specification and other proponents of RDFa joining the WHATWG mailing list to advocate for their point of view, leading to a dispute over RDFa that would continue, off and on, for approximately the next year.

As a “lurker” on the WHATWG mailing list during the period of the dispute, I observed it unfold in real time. I have worked with linked data, RDF, and their predecessor technologies since 1997, and so I am deeply familiar with both linked data and the types of arguments made for and against it. However, for the purposes of this study I did not rely solely on that familiarity but re-immersed myself in the archives of the mailing list in order to attend to the details of how the dispute unfolded in writing.

The WHATWG hosts a public archive of its (no longer active) mailing list, but it is incomplete and not easily searchable. Fortunately the W3C maintains a searchable and complete archive, which I used to carry out my study of the dispute.²³ I began by using the W3C archive’s search functionality to find every message posted to the WHATWG mailing list between August 2008 and July 2009 that contained the terms *RDFa*, *RDF*, *metadata*, or

¹⁹ Ian Hickson, “WHAT open mailing list announcement,” whatwg@whatwg.org mailing list archives, W3C, June 5, 2004, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2004Jun/0000.html>.

²⁰ Matt Bonner, “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 20, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0300.html>.

²¹ Tab Atkins Jr., reply to “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 20, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0301.html>.

²² Ian Hickson, reply to “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 20, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0305.html>.

²³ whatwg@whatwg.org mailing list archives, W3C, <https://lists.w3.org/Archives/Public/public-whatwg-archive/>.

microdata anywhere in the subject line or body of the message. I then identified the thread (tree of messages created by senders replying to earlier messages, typically all having the same subject line) to which each message belonged and read it in its entirety. As I read, I took note of arguments that appealed to concepts beyond the immediate technical details of RDFa implementation. After reading through these threads and reviewing my notes, I identified and organized my notes under five major themes: collaboration and coordination, expertise and consensus, community and scale, human labor and machine labor, and the role of search engines. In my analysis below I interpret these themes in terms of Boltanski and Thévenot’s common worlds model, but I did not initially set out to find these worlds in the RDFa dispute: it was only after the process described above that I began to see the relevance of their model to the themes I had identified.

The Industrial Case against RDFa

The WHATWG discussions I read often revolved around questions of proper process. Ian Hickson (founder of the WHATWG and editor of the HTML specification) repeatedly outlined the WHATWG’s process for evaluating RDFa or any other proposed addition to the HTML specification. First, the problem to be solved by the proposed addition must be clearly stated in the form of a “use case.”²⁴ Next it must be demonstrated that the problem is “is one that needs solving,”²⁵ which can be done by presenting evidence that “the bulk of users”²⁶ face the problem and that current solutions are inadequate “hacks.”²⁷ If the proposed problem passes these tests, the next step is to enumerate all possible solutions in order to identify the one that most effectively and efficiently solves the problem. Finally, “the relevant implementors”—browser vendors, search engine companies, or creators of HTML authoring and validation tools—must show a willingness to implement the identified solution.²⁸ At each stage, participants in the process must “use rational debate, back up their opinions with logical arguments, present research to justify their claims, and

²⁴ Ian Hickson, reply to “RDFa is to structured data, like canvas is to bitmap and SVG is to vector,” [whatwg@whatwg.org mailing list archives, W3C, January 18, 2009, https://lists.w3.org/Archives/Public/public-whatwg-archive/2009Jan/0210.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2009Jan/0210.html).

²⁵ Ian Hickson, reply to “Creative Commons Rights Expression Language,” [whatwg@whatwg.org mailing list archives, W3C, August 21, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0315.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0315.html).

²⁶ Ian Hickson, reply to “Creative Commons Rights Expression Language,” [www-archive@w3.org mailing list archives, W3C, August 24, 2008, https://lists.w3.org/Archives/Public/www-archive/2008Aug/0073.html](https://lists.w3.org/Archives/Public/www-archive/2008Aug/0073.html). Note that this message was posted to a part of the thread that spilled out of the WHATWG mailing list onto another mailing list—a not uncommon occurrence.

²⁷ Hickson, reply to “RDFa is to structured data,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2009Jan/0210.html>.

²⁸ Hickson.

derive proposals from user needs.”²⁹ These arguments were to be evaluated solely “based on their technical merits and on what supporting research they have, and not on the number of times they were made” or the number of people making them.³⁰

According to the criteria outlined above, in the eyes of many members of the WHATWG human-authored metadata on the Web had already failed to pass muster as a reliable and useful technology: “We have shown time and time again that when metadata mechanisms face the wider Web community, they fail.”³¹ The case against human-authored metadata on the Web goes as follows: such metadata only works in “controlled environments” such as “a small coherent community where all the participants have compatible goals.”³² Within these controlled environments, it is possible to ensure that people create and use metadata honestly, conscientiously, and intelligently. But in the absence of mechanisms for enforcing this discipline, some people’s “inherent greed and evilness” will lead them to author metadata dishonestly.³³ Others will be lazy, not caring enough “to bother doing a good job”³⁴ and content to live with “terrible metadata hygiene [*sic*].”³⁵ And even among those people who work diligently and honestly, many will be incompetent, unable to create and use metadata without “making huge mistakes.”³⁶ As a result, outside of controlled environments, “the metadata becomes an utter mess, misused, wrong, missing, syntactically incorrect, semantically incorrect, unusable.”³⁷

The arguments marshaled against human-authored metadata on the Web by WHATWG participants closely followed a template established by influential tech pundit and science

²⁹ Hickson.

³⁰ Ian Hickson, reply to “Self-imposed RDFa cool-down period,” [whatwg@whatwg.org mailing list archives, W3C, August 29, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0604.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0604.html).

³¹ Ian Hickson, “RDFa,” [whatwg@whatwg.org mailing list archives, W3C, August 22, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0345.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0345.html).

³² Hickson.

³³ Ian Hickson, reply to “RDFa Problem Statement (was: Creative Commons Rights Expression Language),” [whatwg@whatwg.org mailing list archives, W3C, August 26, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0416.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0416.html).

³⁴ Ian Hickson, reply to “Creative Commons Rights Expression Language,” [www-archive@w3.org mailing list archives, W3C, August 21, 2008, https://lists.w3.org/Archives/Public/www-archive/2008Aug/0024.html](https://lists.w3.org/Archives/Public/www-archive/2008Aug/0024.html).

³⁵ Hickson, “RDFa,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0345.html>.

³⁶ Ian Hickson, reply to “RDFa,” [whatwg@whatwg.org mailing list archives, W3C, August 22, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0354.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0354.html).

³⁷ Hickson, “RDFa,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0345.html>.

fiction author Cory Doctorow in his 2001 essay “Metacrap.”³⁸ Doctorow began his argument against reliable human-authored metadata by observing that “people lie,” “people are lazy,” and “people are stupid.”³⁹ An implicit assumption of Doctorow’s argument was that the Web is an environment that is uncontrolled and inherently uncontrollable—a tenet among libertarian Web evangelists since its inception.⁴⁰ The uncontrollable nature of the Web means that reliable human-authored metadata is “a pipe-dream,” Doctorow argued, but there is an alternative that works reliably: implicit, observational metadata, derived through statistical analysis of human behavior.⁴¹ The premise that aggregation and analysis of large-scale data could work reliably in uncontrolled environments where human-authored metadata was destined to fail was picked up a few years later by tech pundit and consultant Clay Shirky, who incorporated it into a just-so story purporting to explain the success of Google and the decline of its competitor Yahoo.⁴² Both Shirky and Doctorow advocated addressing the problem of the uncontrolled Web through the carefully controlled creation, analysis and extraction of value from large datasets, derived from the Web but not part of it. Their exemplar was the Google search index: impossible to construct without the Web, but not itself available as a resource openly published on the Web. It would provide a blueprint for today’s machine learning models, trained on data from the open Web but only rarely open to inspection themselves.

Echoing Doctorow and Shirky, RDFa critics on the WHATWG mailing list asserted that the success of web search engine companies obviated the need for human-authored metadata, since “search engines probably already ‘understand’ pages with far more accuracy than most authors will ever be able to express.”⁴³ While acknowledging limits to that understanding, they argued that investment of technical effort should focus on transcending those limits through advances in natural language processing: “Google’s experience is that natural language processing of the actual information seen by the actual end user is far, far more reliable than any source of metadata. Thus from Google’s perspective, investing in

³⁸ Doctorow, “Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia,” August 26, 2001 (archived August 31, 2001 at <https://web.archive.org/web/20010831024354/http://www.well.com/~doctorow/metacrap.htm>).

³⁹ Doctorow, “Metacrap.”

⁴⁰ John Perry Barlow, “A Cyberspace Independence Declaration,” February 9, 1996 (archived December 20, 1996 at https://web.archive.org/web/19961220115636/http://www.eff.org/pub/Publications/John_Perry_Barlow/barlow_0296.declaration).

⁴¹ Doctorow, “Metacrap.”

⁴² Shirky, “Ontology is Overrated: Categories, Links, and Tags,” Clay Shirky’s Writings About the Internet (archived May 18, 2005 at https://web.archive.org/web/20050518032708/http://shirky.com:80/writings/ontology_overrated.html).

⁴³ Hickson, reply to “RDFa Problem Statement,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0416.html>.

RDFa seems like a poorer investment than investing in natural language processing.”⁴⁴ Human authoring of metadata is a poor investment because it “require[s] us to solve a fundamentally unsolvable pair of problems (making humans not be lazy and making humans not be evil).”⁴⁵ Rather than technical resources being squandered on unsolvable problems of human nature, they should be invested in ventures more likely to result in effective functionality: “To scale to the whole Web, the only thing I can see working is the computers understanding human language.”⁴⁶

The arguments made against RDFa and in favor of natural language processing reflect a system of values characteristic of what Boltanski and Thévenot identify as an *industrial* world in which the common good is equated with efficient functioning.⁴⁷ In this world, expert technicians are valued for the responsibility they assume over the planning and execution of realistic projects for the future.⁴⁸ The valorization of technicians is linked to their ability to exercise *control*, not through the direct exercise of power over others but through decomposing complexity into less complex elements and “predicting less complex actions by integrating them into a larger overall plan.”⁴⁹ This decomposition and re-integration is achieved by marshaling lists of use cases and requirements, tools of definition, measurement, and planning, “the instruments for defining and measuring [that] constitute the situation of action as a *problem* leading to the formulation of *hypotheses* and calling for a solution.”⁵⁰ Immorality in the industrial world is associated with laziness and inefficiency due to a lack of motivation or qualifications. These vices lead to waste, pollution, and deterioration—“mess” and “crap”—all symptoms of a poorly controlled system.⁵¹ But harmonious functioning can be restored through the introduction of carefully controlled machinery and “operations of *standardization* and *formalization* [making] it possible to see the

⁴⁴ Ian Hickson, “Trying to work out the problems solved by RDFa,” whatwg@whatwg.org mailing list archives, W3C, December 31, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Dec/0404.html>.

⁴⁵ Hickson, reply to “RDFa Problem Statement,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0416.html>.

⁴⁶ Ian Hickson, reply to “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 28, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0476.html>.

⁴⁷ Boltanski and Thévenot, *On Justification*, 118–123, 203–211.

⁴⁸ Boltanski and Thévenot, 206.

⁴⁹ Boltanski and Thévenot, 209.

⁵⁰ Boltanski and Thévenot, 208.

⁵¹ Boltanski and Thévenot, 205.

world through data *expressed* in numbers, *quantified*, ready to be *processed*, combined, *added up*.”⁵²

The Connectionist Case for RDFa

The arguments made by proponents of RDFa and linked data reflect a system of values characteristic of what Boltanski and Chiapello refer to as a *connectionist* world.⁵³ In this world, project managers are valued for their ability to tap into their personal networks and quickly assemble teams with complementary knowledge and skills.⁵⁴ The status of project managers is linked to their role as redistributors of information and links, putting formerly separated people into contact by integrating them into new networks.⁵⁵ This redistribution of links is achieved through the cultivation of informal relations with trusted partners from past projects. These relations are maintained through savvy use of new communication technologies, so that they may be kept in reserve and re-engaged when necessary for new projects.⁵⁶ Immorality in the connectionist world is associated with an inability to engage in new projects, due either to a failure to communicate effectively or an unwillingness to reciprocally share information and links. These vices lead to the monopolistic hoarding of information within closed networks: silos that benefit privileged insiders but do not extend the network for the greater good.⁵⁷ Hence the importance of lowering barriers to communication and coordination, allowing links to proliferate and ensuring that the network will be able to flexibly support innovative projects.

The dispute over RDFa on the WHATWG mailing list was instigated by a proposal⁵⁸ submitted to the W3C by Creative Commons, a nonprofit founded in 2001 to provide “flexible, customizable intellectual-property licenses” that people posting work on the Web

⁵² Boltanski and Thévenot, 210.

⁵³ Boltanski and Chiapello, *New Spirit*, 103–163.

⁵⁴ Boltanski and Chiapello, 115–116.

⁵⁵ Boltanski and Chiapello, 121–122.

⁵⁶ Boltanski and Chiapello, 117–118.

⁵⁷ Boltanski and Chiapello, 119–120.

⁵⁸ *ccREL: The Creative Commons Rights Expression Language*, ed. Hal Abelson, Ben Adida, Mike Linksvayer, and Nathan Yergler (W3C Member Submission, May 1, 2008), <https://www.w3.org/Submission/2008/SUBM-ccREL-20080501/>.

could use to legally define acceptable uses of that work.⁵⁹ The proposal described the Creative Commons Rights Expression Language (ccREL), a method for embedding these licenses in Web pages using RDFa. As ccREL was one of the first major applications envisioned for RDFa, Creative Commons also became involved in the design of RDFa itself: as the proposal puts it, “RDFa was designed by the W3C with Creative Commons’ input.”⁶⁰ This is relevant because, as the ccREL proposal demonstrates, the Creative Commons vision was deeply rooted in the connectionist world. Boltanski and Chiapello note that the innovative project managers of the connectionist world are modeled on artists and scientists.⁶¹ Creative Commons also centers artists and scientists as subjects: the ccREL proposal concludes by asserting that “Creative Commons wants to make it easy for artists and scientists to build upon the works of others when they choose to.”⁶² Artistic and scientific collaboration is framed as network-building activity to be stimulated through the provision of technological tools and infrastructure that lower barriers to sharing information and making mutual connections: “the minimal infrastructure required to enable collaboration and invention, while letting it flourish as an organic, distributed process.”⁶³

In the artistic-scientific model of collaboration valorized by Creative Commons, small groups work autonomously while maintaining an openness to opportunistic links with other groups. Maintaining local autonomy without foreclosing on opportunities for connection is repeatedly emphasized by advocates for RDFa and linked data: “... there are a very large set of very small problem spaces relevant to a small group at a time. Like RDF itself, RDFa is meeting the problem of allowing these people to share machine-processable data without previously coordinating their approach.”⁶⁴ Linked data exemplifies the close association between connections and information in the connectionist world: “Information is at once the result and the condition of multiplying connections ... To succeed in discovering good connections, such information must be integrated into a representation of the universe to be explored. In a network world, however, there can be no question of an overarching

⁵⁹ Hal Plotkin, “All Hail Creative Commons / Stanford Professor and Author Lawrence Lessig Plans a Legal Insurrection,” *SFGate*, February 11, 2002 (archived November 5, 2012 at <https://web.archive.org/web/20121105090944/http://www.sfgate.com/news/article/All-Hail-Creative-Commons-Stanford-professor-2874018.php>).

⁶⁰ *ccREL*, RDFa and concrete syntax for Work properties, <https://www.w3.org/Submission/2008/SUBM-ccREL-20080501/#SECTION00051000000000000000>.

⁶¹ Boltanski and Chiapello, *New Spirit*, 115–116.

⁶² *ccREL*, Conclusion, <https://www.w3.org/Submission/2008/SUBM-ccREL-20080501/#SECTION00080000000000000000>.

⁶³ *ccREL*, Conclusion.

⁶⁴ Charles McCathieNeville, reply to “Trying to work out the problems solved by RDFa,” whatwg@whatwg.org mailing list archives, W3C, January 1, 2009, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Dec/0416.html>.

representation. Useful representations are local, singular, circumstantial ...”⁶⁵ To cultivate relations of interdependence and trust, people in the connectionist world need “fine-grained, open information” rather than comprehensive standardization.⁶⁶ Fine granularity means that representations can be local but overlapping: “... what RDF is really about is publishing data in a fine-grained enough matter that applications can easily overlap. That's why you can ignore parts of the data if you don't need it. You get a much more loosely-coupled, opportunistic Web, that way ...”⁶⁷

The loose coupling of small autonomous groups is essential to how innovation and creativity are conceived in the connectionist world. Rather than being associated with inspired genius, in the connectionist world “creativity is a function of the number and quality of links. Moreover, it is a matter of *recombination*, rather than creation *ex nihilo*, and readily assumes a ‘distributed’ form (as one talks of ‘distributed intelligence’), with responsibility for innovation being allocated between different actors.”⁶⁸ “Distributed innovation” is a recurrent theme of linked data advocates, who argue that adding support for RDFa to the HTML specification “will help make many small communities happy, each in their own way ... That's the power of RDF, and the idea behind RDFa is to enable that distributed innovation within HTML.”⁶⁹ Making an argument against central planning similar to that made by the Austrian economist Friedrich Hayek⁷⁰—but with fine-grained, open data playing the role of information about prices—linked data advocates argue that an unknown future is best prepared for by enabling distributed intelligence. Creative Commons argues that needs “not yet envisioned” must be addressed through innovation that proceeds “in a distributed fashion in different communities,”⁷¹ just as Hayek argued that “decentralized planning by many separate persons,” by making use of each individual’s “special knowledge

⁶⁵ Boltanski and Chiapello, *New Spirit*, 113.

⁶⁶ Boltanski and Chiapello, 130.

⁶⁷ Ben Adida, reply to “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 29, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0568.html>.

⁶⁸ Boltanski and Chiapello, *New Spirit*, 129.

⁶⁹ Ben Adida, reply to “RDFa,” whatwg@whatwg.org mailing list archives, W3C, August 22, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0366.html>.

⁷⁰ Hayek, “The Use of Knowledge in Society,” *The American Economic Review* 35, no. 4 (September 1945): 519–530.

⁷¹ *ccREL*, Creative Commons and RDF, <https://www.w3.org/Submission/2008/SUBM-ccREL-20080501/#SECTION00021000000000000000>.

of circumstances of the fleeting moment not known to others,” would outperform centralized industrial planning.⁷²

Tim Berners-Lee identified “non-centralisation” as a core requirement for the Web, asserting that it “must allow existing systems to be linked together without requiring any central control or coordination.”⁷³ This specter of a “central bottleneck” is frequently invoked by linked data advocates. For example, in the RDFa dispute, critics of linked data often pointed to microformats as a superior alternative to RDFa. As explained above, microformats are an alternative means of embedding metadata in HTML, the main difference being that microformat metadata vocabularies⁷⁴ are developed through a centralized standardization process very similar to the one WHATWG adopted for the HTML specification.⁷⁵ RDFa advocates repeatedly rejected this process as “unworkable”⁷⁶ for the distribution innovation they envisioned, as it would introduce a point of centralized control:

It seems "bloodtype" is more important in Japanese culture than in Western Europe, but that the toolset and design provided by RDFa allows independent extension of FOAF in Japan without expensive central bottlenecks.⁷⁷

Who decides what the right due diligence is? One organization for *all* topics, ever? An RDF vocabulary can be created by the proper community ... rather than assuming that one central group should be the centralized bottleneck for all development. In other

⁷² Hayek, “Use of Knowledge,” 522.

⁷³ Tim Berners-Lee, “Information Management: A Proposal,” March 1989, <https://www.w3.org/History/1989/proposal.html>.

⁷⁴ A metadata vocabulary is a fixed set of terms for naming the kinds of things to be described, what properties those things may have, and (in some cases) the values that those properties can take. A metadata vocabulary that is designed to support formal reasoning and inference about the described things is known as an “ontology.” Metadata vocabularies defined using RDF support formal reasoning and inference, and this was often cited by advocates as a point in favor of RDFa over alternatives such as microformats. Critics of RDFa, on the other hand, expressed doubts that formal reasoning was either effective or necessary.

⁷⁵ “So You Wanna Develop a New Microformat?” Microformats Wiki, last modified October 6, 2005 (archived October 19, 2005 at <https://web.archive.org/web/20051019000344/http://microformats.org/wiki/process>).

⁷⁶ Ben Adida, reply to “Creative Commons Rights Expression Language,” whatwg@whatwg.org mailing list archives, W3C, August 21, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0331.html>.

⁷⁷ Dan Brickley, reply to “RDFa Problem Statement,” whatwg@whatwg.org mailing list archives, W3C, August 26, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0414.html>.

words, RDF vocabularies function like the web does: decentralized, let the best sites/ vocabs win.⁷⁸

What happens when the people you're justifying your design to are the gatekeepers? What happens when they don't understand the problem you're attempting to solve? Or they disagree with you on a philosophical level? Or they have some sort of political reason to not allow your vocabulary to see the light of day (think large multi-national vs. little guy)?⁷⁹

For similar reasons, linked data advocates also rejected natural language processing and machine learning as potential network-killers. Since these technologies require expensive investments in expertise and computation power that few organizations are in the position to make, relying on them introduces another potential point of central control. RDFa advocates repeatedly expressed concern about making connection and collaboration the Web dependent on Google or other “big tech” intermediaries:

I'm not sure a web design should be predicated on the existence of Google ...⁸⁰

We can reasonably assume the existence of large search engines covering a good part of the public Web. Google being a well known example. But we can't necessarily assume their owners will offer reliable machine-friendly APIs [application programming interfaces] to that data, with terms of service that are sufficiently unconstrained ... the constraints are significant. And may change at any time ...⁸¹

It would seem important that the Web easily enable small-time users of data to efficiently communicate with one another, without the need to have one of the giants as an intermediary ... Google ... can afford to run a huge organisation with massive computer power and many engineers ... there are many others who find that processing structured data is more efficient for their needs than doing free-text analysis of web

⁷⁸ Ben Adida, reply to “Just create a Microformat for it’ - thoughts on micro-data topic,” whatwg@whatwg.org mailing list archives, W3C, May 5, 2009, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0039.html>.

⁷⁹ Manu Sporny, reply to “Just create a Microformat for it’ - thoughts on micro-data topic,” whatwg@whatwg.org mailing list archives, W3C, May 6, 2009, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0045.html>.

⁸⁰ Ben Adida, reply to “RDFa,” whatwg@whatwg.org mailing list archives, W3C, August 22, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0370.html>.

⁸¹ Dan Brickley, reply to “RDFa,” whatwg@whatwg.org mailing list archives, W3C, August 23, 2008, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0375.html>.

pages ... these are the people who [*sic*] have decided that investing in RDFa is a far more valuable exercise [*sic*] than trying to out-invest Google in natural language processing.⁸²

These concerns reflect the characteristic moral sense of the connectionist world, which condemns “the networker [who] keeps information to himself, weaves connections in secret ... avoiding a situation where others can pursue them without going through him ... [whose] monopolistic conduct leads fairly rapidly to the closure of the network” such that it no longer serves the common good.⁸³

Worlds in Conflict

Conflicting visions of the common good, not disagreement over technical issues, precipitated the dispute over RDFa. Technical disagreements can be resolved through an industrial test of strength such as the process prescribed by the WHATWG. But such a resolution is impossible when “the very reality of the common good underlying the legitimacy of the test is contested,” as it was in this case.⁸⁴ The WHATWG appealed to RDFa advocates to submit to an industrial test, insisting that “it is important to actually make sure the problem you are solving is one that needs solving,”⁸⁵ but RDFa advocates denounced the test as invalid: “Who gets to decide which problems need solving?”⁸⁶ Having to justify RDFa by demonstrating that it would solve a problem for “the bulk of users” ran counter to the fine granularity linked data advocates required to enable distributed innovation: “Are we only trying to solve problems that the *bulk* of users know they have? What about enabling new solutions that will provide a new category of solutions that the bulk of users can't quite put their finger on?”⁸⁷ Linked data advocates' focus on small communities made no sense to the WHATWG: “There are thousands of small communities with their own needs, we can't possibly address each one in HTML. Indeed, we have design

⁸² McCathieNeville, reply to “Trying to work out,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Dec/0416.html>.

⁸³ Boltanski and Chiapello, *New Spirit*, 120.

⁸⁴ Boltanski and Thévenot, *On Justification*, 223.

⁸⁵ Hickson, reply to “Creative Commons,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0315.html>.

⁸⁶ Adida, reply to “Creative Commons,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0331.html>.

⁸⁷ Ben Adida, reply to “Creative Commons Rights Expression Language,” www-archive@w3.org mailing list archives, W3C, August 24, 2008, <https://lists.w3.org/Archives/Public/www-archive/2008Aug/0076.html>.

principles that make addressing the needs of small communities an explicit non-goal.”⁸⁸ Those principles, linked data advocates countered, could not be applied to metadata vocabulary design: “A somewhat strained analogy would be bringing in representatives from all of the cultures of the world and having them agree on a universal vocabulary. It is an untenable prospect, there is too much diversity in the world to agree on one master vocabulary.”⁸⁹ In the connectionist world, a universal solution is untenable, but in the industrial world it is the very definition of the common good: “That's pretty much exactly what Unicode did. Or what we're doing with HTML. That doesn't seem untenable [*sic*], it seems quite reasonable.”⁹⁰

RDFa advocates' disagreement with the WHATWG exemplifies how people in the connectionist world call the industrial world into question:

... flexibility, their ability to adapt and learn continuously, become major advantages, which take precedence over their technical expertise (knowledge changes so quickly) and their experience. Personality make-up, the qualities of communication, listening and openness to differences, thus count for more than efficiency as measured by the ability to achieve predefined objectives. Work methods are developed in line with constantly changing needs: people organize themselves and invent local rules that are not amenable to totalization and comprehensive rationalization by some putative organization department.⁹¹

Ultimately this disagreement was resolved through an uneasy compromise. Of the use cases identified for RDFa, only two were deemed efficiently solvable through changes to the HTML specification: providing metadata to search engines that could be used to enhance the presentation of search results⁹² and “annotat[ing] structured data that HTML has no semantics for, and which nobody has annotated before, and may never again, for private use

⁸⁸ Hickson, “RDFa,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0345.html>.

⁸⁹ Manu Sporny, “RDFa Problem Statement (was: Creative Commons Rights Expression Language),” [whatwg@whatwg.org mailing list archives, W3C, August 25, 2008, https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0410.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0410.html).

⁹⁰ Hickson, reply to “RDFa Problem Statement,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2008Aug/0416.html>.

⁹¹ Boltanski and Chiapello, *New Spirit*, 135.

⁹² Ian Hickson, “Providing enhanced search results,” [whatwg@whatwg.org mailing list archives, W3C, May 19, 2009, https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0269.html](https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0269.html).

or use in a small self-contained community.”⁹³ The latter was an attempt to reformulate the values of the connectionist world in the form of a “problem” acceptable to the industrial world—part of the process of working out a compromise between two worlds.⁹⁴

RDFa itself, however, was rejected as a solution to these problems in favor of an alternative, dubbed “microdata,” designed by Ian Hickson to better agree with his engineering sensibilities. Microdata was simpler and less flexible than RDFa, making it easier to write efficient microdata parsers. But from the perspective of the connectionists, the most significant difference was that microdata rejected the use of URLs as identifiers, thus eliminating the one feature of RDF that promotes visions of a globally distributed network of data. Hickson made it clear that microdata was designed not to enable the construction of this network but instead “for private use or use in a small self-contained community.”⁹⁵

Two years later, microdata would become the initial focus of schema.org, a consortium founded by Google, Microsoft, and Yahoo to manage one master vocabulary that website administrators could use to “help search engines and other applications better understand your content and display it in a useful, relevant way.”⁹⁶ Schema.org was a response to what became the first major use of RDFa: Facebook’s Open Graph Protocol, intended to ease the process of incorporating Web content into Facebook’s “social graph.”⁹⁷ Ironically, rather than devolving power to small independent communities, embedding linked data in Web pages accelerated the centralized hoarding of information by providing a way for “content producers” to make themselves more legible to these platforms’ data processing regimes. In retrospect, linked data advocates undermined their own vision by focusing on technical standards for encoding metadata rather than the much harder problem of establishing a framework of general conventions within which independent groups could define their own vocabularies in a decentralized-yet-loosely-connectable way. As a result, organizations often choose to use linked data not in order to exercise local autonomy, but in exchange for preferential treatment on centralized platforms, to whom they relinquish control over

⁹³ Ian Hickson, “Annotating structured data that HTML has no semantics for,” whatwg@whatwg.org mailing list archives, W3C, May 10, 2009, <https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0116.html>.

⁹⁴ Boltanski and Thévenot, *On Justification*, 281.

⁹⁵ Hickson, “Annotating structured data,” <https://lists.w3.org/Archives/Public/public-whatwg-archive/2009May/0116.html>.

⁹⁶ “Why Use Microdata?” Getting Started with Schema.org, Schema.org (archived on June 6, 2011 at https://web.archive.org/web/20110606061323/http://www.schema.org/docs/gs.html#microdata_why).

⁹⁷ “The Open Graph Protocol,” last modified December 10, 2010 (archived on December 11, 2010 at <https://web.archive.org/web/20101211091635/http://ogp.me/>).

vocabulary design decisions rather than investing in the labor and expertise necessary to retain control themselves.⁹⁸

By making these observations, I am not suggesting that practitioners considering the use of linked data should necessarily adopt connectionist values. Instead I am emphasizing that linked data is not a neutral technology: it comes equipped with its own moral logic, a logic that may not sit easily with the values espoused by organizations implementing it. For example, linked data and the Web are sometimes assumed to be “democratizing” forces well-suited to the aims of civic institutions like libraries and local governments. But the common good understood as a collective good, realizable through democratic expression of the general will, is not how the common good is understood in the connectionist world.⁹⁹ Nor does the connectionist world, despite its reliance on the maintenance of local connections, respect communitarian rootedness, traditional ways of knowing, or the authoritative wisdom of elders and ancestors—calling into question the idea that linked data is particularly well suited to organizing indigenous cultural heritage or projects of decolonization.¹⁰⁰ Boltanski and Chiapello argue that, beginning in the 1970s, critiques of capitalist institutions were successfully defused through the deployment of a connectionist grammar that elides the differences between putting on a play, organizing a protest, and closing a factory: all become “projects.”¹⁰¹ The appeal of linked data may be due in part to the ease with which the connectionist grammar can assimilate such a variety of activities.

On the other hand, it may be that linked data advocates do in fact believe in the connectionist ideals that Creative Commons and others have made seem so attractive. If this is the case, then there is still a need for clearer articulation of, and mechanisms for enforcing, the principles of justice native to the connectionist world—principles that Boltanski and Chiapello refer to as the “projective city.” Boltanski and Chiapello identify three kinds of ways that unjust exploitation in the connectionist world might be regulated:

1. frameworks for inventorying contributions to the network so as “to put an end to exploitation that is bound up with the low visibility of certain contributors”;¹⁰²

⁹⁸ Erik Radio and James Kalwara, “The Trajectory of Linked Data in Late Capitalism,” *Journal of Documentation* 78, no. 3 (2022): 597–612, <https://doi.org/10.1108/JD-02-2021-0037>.

⁹⁹ Boltanski and Thévenot (On Justification, 107–117, 185–193) name the *civic* world as the one that identifies the common good as democratic expression of a collective will. For how connectionist goods can be mistaken for civic ones, see Alejandro Diaz, “Through the Google Goggles: Sociopolitical Bias in Search Engine Design,” in *Web Search: Multidisciplinary Perspectives*, ed. Amanda Spink and Michael Zimmer (Berlin: Springer-Verlag, 2008), 11–34.

¹⁰⁰ Boltanski and Chiapello, *New Spirit*, 133–135.

¹⁰¹ Boltanski and Chiapello, 111.

¹⁰² Boltanski and Chiapello, 382.

2. fairer rules of remuneration, modeled after the collective agreements pursued by unionized workers, but where “remuneration” is understood not only in terms of fairer pay for participation in individual projects, but also as fairer opportunities to acquire and maintain the skills and reputation necessary to ensure participation in future projects;¹⁰³ and
3. equality of opportunity for mobility, through mechanisms intended to provide everyone (not only those who are already well-integrated into the network) with opportunities to cultivate new links and reestablish those that have withered.¹⁰⁴

Their analysis suggests that linked data advocates should focus on

1. standards and methods for comprehensively crediting contributions to the network of linked data projects, no matter how small;
2. rules for providing contributors with rewards in forms besides public recognition or the warm feelings yielded by volunteering; and
3. programs for ensuring that not only the already-well-connected (e.g. the Wikimedia Foundation, Yale University, or the Getty Research Institute) have the opportunity to grow their networks.

Alongside these efforts (and particularly in relation to #2 above) there needs to be recognition and renewed critique of how linked data, by facilitating the process of analyzing and codifying resources into distinctive but recombinable features,¹⁰⁵ potentially renders all kinds of resources newly amenable to commodification: “ideas ... and information about other people’s [social] relations or ... their state of health, their political, aesthetic, intellectual, etc., inclinations” and so on.¹⁰⁶

Toward Diplomacy among Reflexive Technical Institutions

Distinct and incompatible visions of the common good led to the dispute over RDFa and prevented its resolution through the WHATWG specification process. That process allowed for the evaluation of competing designs with respect to an industrial ideal of efficient and effective functioning, but it did not allow for consideration of any competing ideals. Hence the frustration of RDFa advocates, who sought in vain to promote their ideal

¹⁰³ Boltanski and Chiapello, 384–391.

¹⁰⁴ Boltanski and Chiapello, 392–398.

¹⁰⁵ Boltanski and Chiapello, 445–446.

¹⁰⁶ Boltanski and Chiapello, 378.

of greater network connectivity in solely technical terms. Rather than repeat that mistake, proponents of linked data might reflect on how their rhetoric reveals their own taken-for-granted assumptions about the common good. If upon reflection they decide to re-confirm their commitment to connectionist ideals, then they can address the challenge of enunciating and advocating for a connectionist ethics. Alternatively, they might conclude that their conception of the common good requires facilities not afforded by linked data and seek alternative ways of building the world they envision.

The ethical use or non-use of linked data cannot be assured by formulating rules of conduct without clarifying the moral principles underlying those rules. Even once those moral principles are clarified, ethical considerations cannot be treated as independent from the technical evaluation of possible architectures for arranging and relating people, technologies, protocols, expressions, and ideas—among which linked data is just one possibility. Deciding how and how not to connect and arrange things is not a separate activity to be carried out before or after moral questions have been addressed; it is *how* moral questions are addressed. Ethically deploying or dismantling technological systems such as linked data depends not on finding the right moral framework beforehand, but on building better technical institutions. Good technical institutions are *reflexive*, attempting as best they can to provide stability and continuity according to some specific conception of the common good. But they are also aware of that conception's limitations and open to the possibility of transformation in response to critiques born of other conceptions.¹⁰⁷ This implies that good technical institutions need other good technical institutions willing to critique, confront and occasionally compromise with them. Perhaps we need to redirect effort away from developing yet more *ethical* frameworks for technology, and toward developing more *diplomatic* ones.

¹⁰⁷ Ryan Shaw, "The Missing Profession: Towards an Institution of Critical Technical Practice," in *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia, June 16-19, 2019, Information Research* 24, no. 4 (2019), <http://informationr.net/ir/24-4/colis/colis1904.html>.

Bibliography

- Bénatouïl, Thomas. "A Tale of Two Sociologies: The Critical and the Pragmatic Stance in Contemporary French Sociology." *European Journal of Social Theory* 2, no. 3 (1999): 379–96. <https://doi.org/10.1177/136843199002003011>.
- Bachman, Charles W. "Data Structure Diagrams." *DATA BASE: A Quarterly Newsletter of the Special Interest Group on Business Data Processing* 1, no. 2 (1969): 4–10. <https://doi.org/10.1145/1017466.1017467>.
- Barlow, John Perry. "A Cyberspace Independence Declaration." February 9, 1996. Archived December 20, 1996 at https://web.archive.org/web/19961220115636/http://www.eff.org/pub/Publications/John_Perry_Barlow/barlow_0296.declaration.
- Berners-Lee, Tim. "Information Management: A Proposal." March 1989, May 1990. <https://www.w3.org/History/1989/proposal.html>.
- Berners-Lee, Tim. "Linked Data." Design Issues. Last modified October 25, 2006. Archived November 15, 2006 at <https://web.archive.org/web/20061115043657/https://www.w3.org/DesignIssues/LinkedData.html>.
- Boltanski, Luc. "A Journey Through French-Style Critique." In *New Spirits of Capitalism? Crises, Justification, and Dynamics*, edited by Paul du Gay and Glenn Morgan, 43–59. Oxford: Oxford University Press, 2013. <https://doi.org/10.1093/acprof:oso/9780199595341.003.0002>.
- Boltanski, Luc. *On Critique: A Sociology of Emancipation*. Translated by Gregory Elliott. Cambridge: Polity Press, 2011.
- Boltanski, Luc and Ève Chiapello. *The New Spirit of Capitalism*. Translated by Gregory Elliott. London: Verso, 2018.
- Boltanski, Luc and Laurent Thévenot. *On Justification: Economies of Worth*. Translated by Catherine Porter. Princeton: Princeton University Press, 2006.
- Diaz, Alejandro. "Through the Google Goggles: Sociopolitical Bias in Search Engine Design." In *Web Search: Multidisciplinary Perspectives*, edited by Amanda Spink and Michael Zimmer, 11–34. Berlin: Springer-Verlag, 2008.
- Doctorow, Cory. "Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia." August 26, 2001. Archived August 31, 2001 at <https://web.archive.org/web/20010831024354/http://www.well.com/~doctorow/metacrap.htm>.
- Hayek, Friedrich A. "The Use of Knowledge in Society." *The American Economic Review* 35, no. 4 (September 1945): 519–530.

- Hogan, Aidan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, et al. *Knowledge Graphs*. Morgan & Claypool, 2022. <https://doi.org/10.2200/S01125ED1V01Y202109DSK022>.
- Larriba-Pey, Josep Lluís, Norbert Martínez-Bazán, and David Domínguez-Sal. "Introduction to Graph Databases." In *Reasoning Web: Reasoning on the Web in the Big Data Era*, edited by Manolis Koubarakis, Giorgos Stamou, Giorgos Stoilos, Ian Horrocks, Phokion Kolaitis, Georg Lausen, and Gerhard Weikum, 171-194. Cham: Springer, 2014. https://doi.org/10.1007/978-3-319-10587-1_4.
- Pepper, Steve. "Topic Maps." In *Encyclopedia of Library and Information Sciences*, edited by John D. McDonald and Michael Levine-Clark, 4th ed., 4611-4623. Boca Raton: CRC Press, 2018.
- Plotkin, Hal. "All Hail Creative Commons / Stanford Professor and Author Lawrence Lessig Plans a Legal Insurrection." *SFGate*, February 11, 2002. Archived November 5, 2012 at <https://web.archive.org/web/20121105090944/http://www.sfgate.com/news/article/All-Hail-Creative-Commons-Stanford-professor-2874018.php>.
- Radio, Erik and James Kalwara. "The Trajectory of Linked Data in Late Capitalism." *Journal of Documentation* 78, no. 3 (2022): 597-612. <https://doi.org/10.1108/JD-02-2021-0037>.
- Shaw, Ryan. "The Missing Profession: Towards an Institution of Critical Technical Practice." In *Proceedings of CoLIS, the Tenth International Conference on Conceptions of Library and Information Science, Ljubljana, Slovenia, June 16-19, 2019*. *Information Research* 24, no. 4 (2019). <http://informationr.net/ir/24-4/colis/colis1904.html>.
- Shirky, Clay. "Ontology is Overrated: Categories, Links, and Tags." Clay Shirky's Writings About the Internet. Archived May 18, 2005 at https://web.archive.org/web/20050518032708/http://shirky.com:80/writings/ontology_overrated.html.
- World Wide Web Consortium (W3C). *ccREL: The Creative Commons Rights Expression Language*. Edited by Hal Abelson, Ben Adida, Mike Linksvayer, and Nathan Yergler. W3C Member Submission, May 1, 2008. <https://www.w3.org/Submission/2008/SUBM-ccREL-20080501/>.
- World Wide Web Consortium (W3C). *Compound Document Use Cases and Requirements Version 2.0*. Edited by Steve Speicher and Petri Vuorimaa. W3C Working Draft, December 19, 2005. <https://www.w3.org/TR/2005/WD-CDFReqs-20051219/>.
- World Wide Web Consortium (W3C). *RDF in XHTML*. Edited by Ben Adida. W3C Task Force Document, October 12, 2004. <https://www.w3.org/2001/sw/BestPractices/HTML/2004-10-12-tf>.
- World Wide Web Consortium (W3C). *RDFa in XHTML: Syntax and Processing*. Edited by Ben Adida, Mark Birbeck, Shane McCarron, and Steven Pemberton. W3C Working Draft, October 18, 2007. <https://www.w3.org/TR/2007/WD-rdfa-syntax-20071018/>.

World Wide Web Consortium (W3C). *Resource Description Framework (RDF) Model and Syntax Specification*. Edited by Ora Lassila and Ralph R. Swick. W3C Recommendation, February 22, 1999. <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.

World Wide Web Consortium (W3C). *W3C TAG Ethical Web Principles*. Edited by Daniel Appelquist and Hadley Beeman. W3C TAG Finding, October 6, 2021. <https://www.w3.org/2001/tag/doc/ethical-web-principles-20211006>.