

特集「Linked Data とオントロジー」にあたって

長野 伸一

(株式会社東芝研究開発センター)

山口 高平

(慶應義塾大学)

セマンティック Web が提唱され 10 年以上が経過したが、依然、研究開発段階に留まっている。当初は大きな関心が寄せられ、RDF データ、RDF スキーマ、オントロジー、ルールの標準化は進んだ。しかしながら、Web コンテンツにメタデータが付与され、エージェントがオントロジーを使って Web コンテンツの意味を理解することにより、さまざまな知的な処理が実現されることが期待されたが、その期待は実現されていない。標準語彙を含むオントロジーはかなり開発されたが、Web コンテンツへのメタデータ付与が進まないことが最大の原因であろう。近年、RDFa、microformat などにより HTML コンテンツを構造化する環境が整ってきたが、それでも、Web ページ提供者側に、既存の HTML コンテンツに構造を付与しようというインセンティブが働かない。B(enefit)byC(ost) が小さいのである。

このような状況下で、2006 年に、Tim Berners-Lee 卿により、セマンティック Web の新しい考え方である Linked Data が提唱され、2007 年 5 月に RDF トリプルが 5 億、RDF リンクが 12 万程度であったものが、2011 年 9 月には、RDF トリプルが 310 億、RDF リンクが 5 億にまでスケールアップしてきた。欧米の行政機関、生命科学、出版、メディア、地理などの分野で Linked Data の公開が進んでいる。特に、米国と英国の行政機関が、PDF や HTML 形式で公開していたデータを Linked Data 形式で再公開することを積極的に推進し、人が見て理解するためのデータから、アプリケーションで加工処理されるデータに移行した結果、さまざまなアプリケーションが試作可能となっている。本特集では、Linked Data の公開方法、共有方法、構築方法、オントロジー、公開ツール、Linked Data を利用したサービス、センサデータなど、幅広い観点から、Linked Data とオントロジーについて解説する。

Linked Data が社会に広く浸透するには、大きなコストをかけることなく、継続的なデータの公開が重要である。神崎正英氏の「Linked Data とデータマッピング」では、我々が日常よく利用する表形式データから RDF データへの変換方法について、ツールを使いながら具体的に説明していただいている。また、Linked Data がアプリケーションで利用されるために、SPARQL や JSON についても言及されている。最後に、小さなデータセットもリンクすることにより新しい価値が生まれるという、重要な指摘がなされている。

Linked Data は、Web の仕組みを利用して、個々の情報や、情報間の関係に URI という識別子を与えて提供さ

れる。武田英明氏の「Linked Data とアイデンティティ」では、Web 上で個々の情報の信頼性を高める手段として認知されつつあるアイデンティティについて、情報の存在に関する形式的な洞察から入り、Web で流通する識別子、語彙/スキーマについて議論されている。また、美術館・博物館の情報共有を目指した LODAC プロジェクトが解説されている。アイデンティティの信頼性、維持、対応付けなど、今後研究すべき課題が提示されている。

Linked Data の作成プロセスに関して、既存データからの変換と並んで特徴的な点は、個人やコミュニティの活動を通してボトムアップにデータを整備できる点にある。江渡浩一郎氏と濱崎雅弘氏の「集合知による Linked Data の構築」では、集合知の一般的議論の後、共同作業により機械可読データを開発する DataWiki の解説に入り、既存 DataWiki として、Freebase、SemanticMediaWiki、OntoWiki が紹介され、著者らの DataWiki として、Social Infobox と Wedata が詳しく解説されている。各 DataWiki ではデータ入力の動機付けが異なっており、これらを比較することは興味深い。

世界規模のデータネットワークを形成しつつある Linked Data を利用して、組織や領域を横断した情報統合の実現性が高まっている。森田武史氏と山口高平の「Linked Data を利用した情報統合」では、情報統合の基盤である RDF ストアと推論エンジン、標準語彙について解説されている。また、日本語 Linked Data を利用した情報統合に関する事例として、地理・位置データを中心とした日本語 Linked Data の構築方法、およびそれらを利用したモビリティサービスが紹介されている。本サービスにより、Linked Data の構築と利用についての全体の開発工程が理解できよう。

長野伸一の「Linked Data とセンサネットワーク」では、今後、ビッグデータの中核をなすセンサデータ分野において、Linked Sensor Data、センサデータ用のオントロジー、活用事例について解説されている。今後、センサデータの重要性はますます高くなるので、このような先駆的な研究開発の紹介は参考になる。

Linked Data は、世界的には順調に普及しているが、日本では、一部の機関からしか公開されておらず、今後の普及が期待される場所である。幸い、本特集の発行と同年月に「Linked Open Data チャレンジ Japan 2011」コンテスト授賞式の開催が予定されており、計 73 件の作品がエントリーされ、我が国でも Linked Data の関心が高まってきた。本特集が、さらなる関心の高まりの一助になれば幸いである。