



# Making W3C better for Web Data Standardization (*practices & tooling*)

Dave Raggett [dsr@w3.org](mailto:dsr@w3.org)

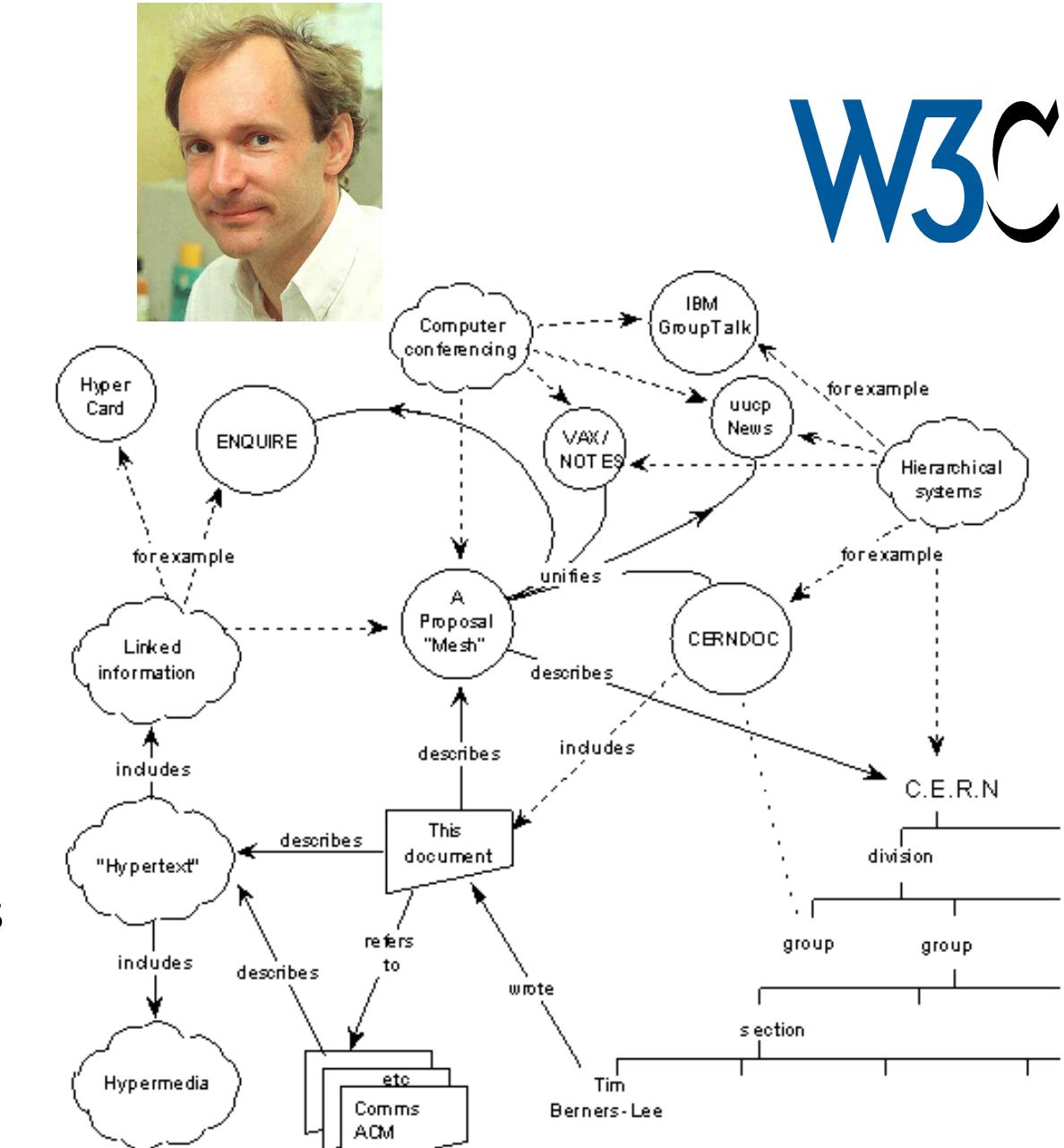
*W3C Web Data Activity Lead*

TPAC 2017

# W3C and Linked Data



- TimBL's original proposal for the Web included an example of a semantic network
  - A binary graph of concepts and relationships
- Resource Description Framework
  - URLs for concepts and relationships
  - Triples (subject, predicate, object)
  - Dereference URL to get further triples
- Followed by a suite of standards
  - RDF Core, RDF Schema, OWL, SPARQL, Turtle, LDP, SHACL, ...



[From Tim's 1990 proposal for the Web](#)

# Web Data



- Covers a broad range of opportunities
  - Annotations to web pages
  - Downloadable data sets
  - Network APIs
  - Open markets of data and services
  - Discovery and semantic interoperability
- Multiple formats and encodings
  - JSON, XML, Turtle, JSON-LD, RDF-a, Microdata and many others
- Interoperability depends on Web Data standards
- ***How can we make W3C a better venue for Web Data standardisation?***

# Open Data



- Data that anyone can access, use or share
  - When big companies or governments release non-personal data, it enables small businesses, citizens and medical researchers to develop resources which make crucial improvements to their communities
  - According to McKinsey, a global market powered by open data from across seven sectors would create between \$3tn and \$5tn a year
  - Open standards are key to enabling markets of data and services based upon open data. Further details are provided on the Open Data Institute page on [what is open data](#).

# Data as a Service



- In some cases, access to data is restricted by agreements between providers and consumers of data and services, possibly involving some form of remuneration along with terms and conditions
- Open markets for such cases are facilitated by open standards for data. These standards can cover data formats and data models as well as other metadata, protocols and APIs
- Some related topics include what standards are needed to support discovery, and whether data is available as datasets for download, or for access via network APIs
- Other topics relate to standards for describing data in terms of shared vocabularies, e.g. units of measure, support for internationalization, and whether the vocabularies are stable or are expected to track rapidly evolving needs



# Current Work

# Linked Data Vocabularies



- W3C Recommendations
  - The Organization Ontology
  - RDF Data Cube Vocabulary
  - Data Catalog (DCAT) Vocabulary
  - PROV-O: The PROV Ontology
  - Time Ontology in OWL
  - Semantic Sensor Network ontology (SSN)
- Other work that is at an earlier stage of maturity
  - vCard Ontology – for describing people and organizations
  - Basic Geo Vocabulary
  - Geospatial Vocabulary
  - Term-centric Semantic Web Vocabulary Annotations

# Best Practices Guidelines



- Best practice guidelines
  - Data on the Web Best Practices when publishing or consuming data on the Web
  - Spatial Data on the Web Best Practices in collaboration with the OGC
  - Best Practice Recipes for Publishing RDF Vocabularies
  - Best Practices for Multilingual Linked Open Data
- This includes consideration of challenges relating to metadata, the data license, provenance and quality, data versioning, data identification, data formats, data vocabularies, data access, data preservation, feedback, data enrichment and data republication.

# W3C Community Groups



- There are many W3C Community Groups with an interest in Open Data Standards

Age Labels Model, Automotive Ontology, Best Practices for Multilingual Linked Open Data, Bioschemas for life sciences, Credentials, CSV on the Web, Data on the Web Best Practices, Data Visualisation, DataSheets, Declarative Linked Data Apps, Digital Asset Management Industry Business Ontology, Digital Verification, Electronic Governance (eGov), Exploration of Scientific Data, Exposing and Linking Cultural Heritage Data, Exposing IEEE LOM metadata as Linked Data, Financial Industry Business Ontology, Geospatial Semantic Web, Healthcare Schema Vocabulary, Human Services, JSON for Linking Data, LDP Next, Linked Building Data, Linked Data for Language Technology, Linked Data Models for Emotion and Sentiment Analysis, Locations and Addresses, Machine Learning Schema, Meat Products, Natural Language Interfaces for the Web of Data, Networked Data, ODRL, Ontology Lexica, Open Annotation, Open Data Directory, Open Data Spain, Open Educational Resources Schema, Open Government, Open knowledge-driven service-oriented system architectures and APIs, Open Linked Education, Open Science, OpenActive, OpenTrack, Organisation Profile Documents, OWL: Experiences and Directions, PDF and Open Data, Permanent Identifier, Places, Property Graphs Model and API, RDF and XML Interoperability, RDF JavaScript Libraries, RDF Stream Processing, Research Object for Scholarly Communication, Restaurant Ontology, Schema Archetypes, Schema Bib Extend, Schema Course extension, Schema Generator, Schema.org, Schema.org for dataset, Scholarly HTML, SDShare, Semantic Building Data, Semantic News, Semantic Open Data, Semantic Sensor Networks, Semantic Statistics, Semantic Water Interoperability Model, Semantic Web in Healthcare and Life Sciences, SHACL, Shape Expressions, SKOS and OWL for Interoperability, Smart Manufacturing, Sport Schema, Traffic Event Ontology, Web Observatory

# Web search and schema.org



- **Success story for metadata on the Web**
- Schema.org founded by Google, Microsoft, Yahoo and Yandex
- Benefits of rich search results from embedded metadata
  - Web pages, email and beyond
  - Businesses incentivized to annotate their websites
  - Used by over 10 million websites
- Vocabularies can be used with a variety of encodings including RDFa, Microdata and JSON-LD
- Vocabularies developed in W3C Community Groups using the [public-schemaorg@w3.org](mailto:public-schemaorg@w3.org) mailing list and through [GitHub](https://github.com/schemaorg)
  - See <https://www.w3.org/community/schemaorg/>

# Web Data Standardization



- How can we make W3C a better venue for Web Data standardisation?
- Short term study with support from the Open Data Institute and InnovateUK
  - Online questionnaire
  - One on one interviews
  - Analysis of W3C server logs
- Report due in mid-December
- See W3C 30 October 2017 blog post announcing the questionnaire
  - <https://www.w3.org/blog/2017/10/questionnaire-on-practices-tooling-for-web-data-standardisation/>

# Questionnaire



- This [questionnaire](#) is open to anyone interested in Web data standardisation, whether you were previously or currently involved in standards group, or have an interest in contributing to future standards
- We plan to make the results of this study public, but will not quote or attribute answers without your permission
- We may contact you to ask follow up questions based upon the information you have provided. We will provide anonymity in respect to who contributed to this study, except when you agree for us to either quote you or seek testimonials from you. We will not share your contact details without your permission
- It would be very much appreciated if you can respond to the [questionnaire](#) by late November 2017
- If you have any questions about the study and this questionnaire, please contact Dave Raggett <[dsr@w3.org](mailto:dsr@w3.org)>, W3C Data Activity Lead.

# Questionnaire



- Divided into several sections
  - About you
  - What kind of data standards?
  - Sustainability and governance
  - Scaling challenges
  - Tooling and practices
  - Liaisons, outreach and community building
  - Miscellaneous feedback on W3C Groups
- Show the questionnaire
  - <https://goo.gl/forms/qiQ5smQ7U1qTl0fp1>
- Invite feedback/discussion on each area