Toward Standards for NoSQL

Sandro Hawke, W3C

NoSQL Live … from Boston
March 11, 2010

Overview

- Three Parts:

    1. Does NoSQL Need More Standards?

    2. How Standards are made at W3C

    3. NoSQL, meet some RDF Standards

The Question

- What new Standards might NoSQL need?

- Currently using:

  - Unicode (from Unicode Consortium)

  - HTTP (from ietf/w3c)

  - HTML (from w3c)

  - XML (from w3c)

  - JSON (from ietf)

  - URI (from ietf/w3c)

  - ...?   Thrift?

What about SQL?

- Maybe a modified SQL?
    - Relaxing ACID?
    - Improved Portability?
    - Improved/HTTP network API?
- How many Concerns would that address?
    - Probably not CouchDB, MongoDB
- "Cultural" Issues

Role of Open Source

Do we still need standards organizations?

- Long ago, ANSI brought order to C

- But we don't need that for Python, Scala, Erlang

- Is anyone worried about vendor lock-in with Cassandra? (I doubt it.)

- But still: DOM, html5, unicode....

There are lots of ways to achieve standards.
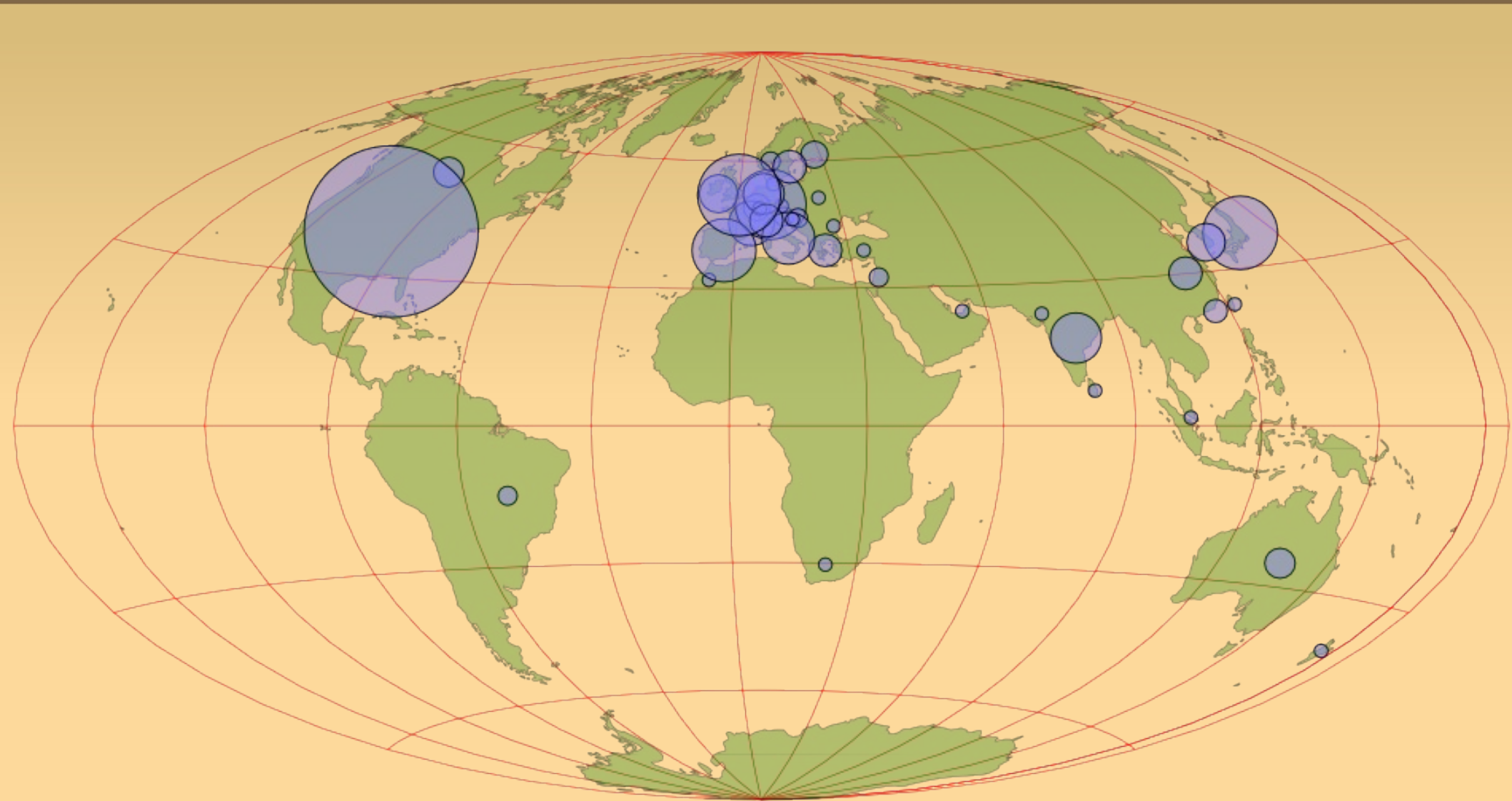
Some Existing Standards...

- Web Architecture (REST)

- XML Schema Datatypes (eg dates)

- I18n / tags for natural languages (BCP 47)

- Semantic Web / Linked Data / RDF

About W3C

- Founded 1994 by Web inventor, Tim Berners-Lee

- Mission: Lead the Web to its Full Potential

    (making the web do more for everyone)

- Funded by membership fees, donations, grants

International Presence

333 member organizations in 40 countries. Offices in 20 countries, including 3 host sites (MIT/US, ERCIM/France, Keio/Japan).

Diverse Membership

Broad Reach

- Translations of standards into 45+ languages

- Liaisons w/ 40+ global standards organizations UN (IGF), ISO, ITU, IETF, OGF, Unicode, Eclipse, OMA, etc.

- 32,000 people subscribed to mailing lists

- 10,000,000 hits/day on www.w3.org

- 100+ Web standards:HTML, XML, Voice, accessibility, etc.

- 1,500+ participants in 60+ Groups

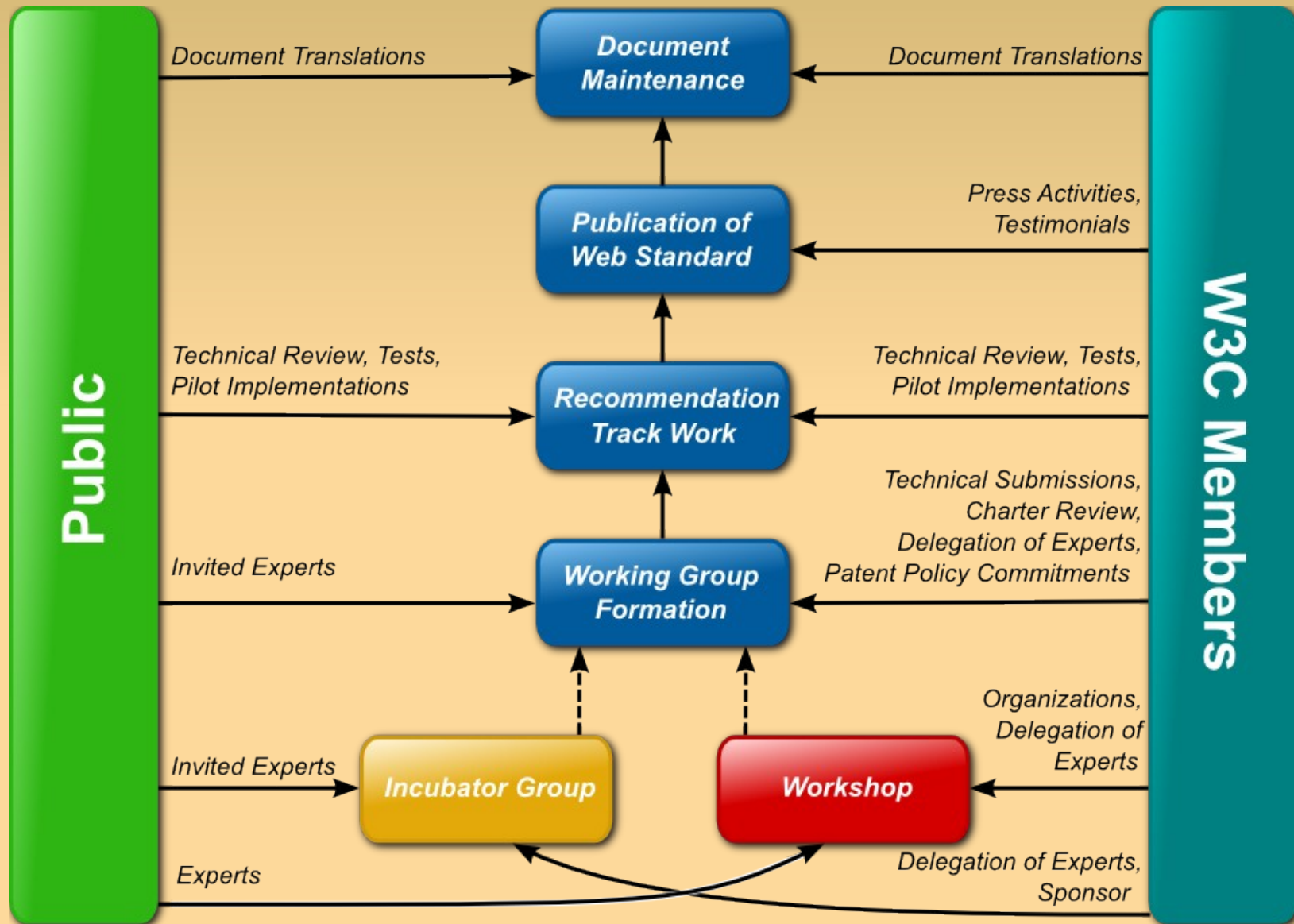W3C WORLD WIDE WEB consortium
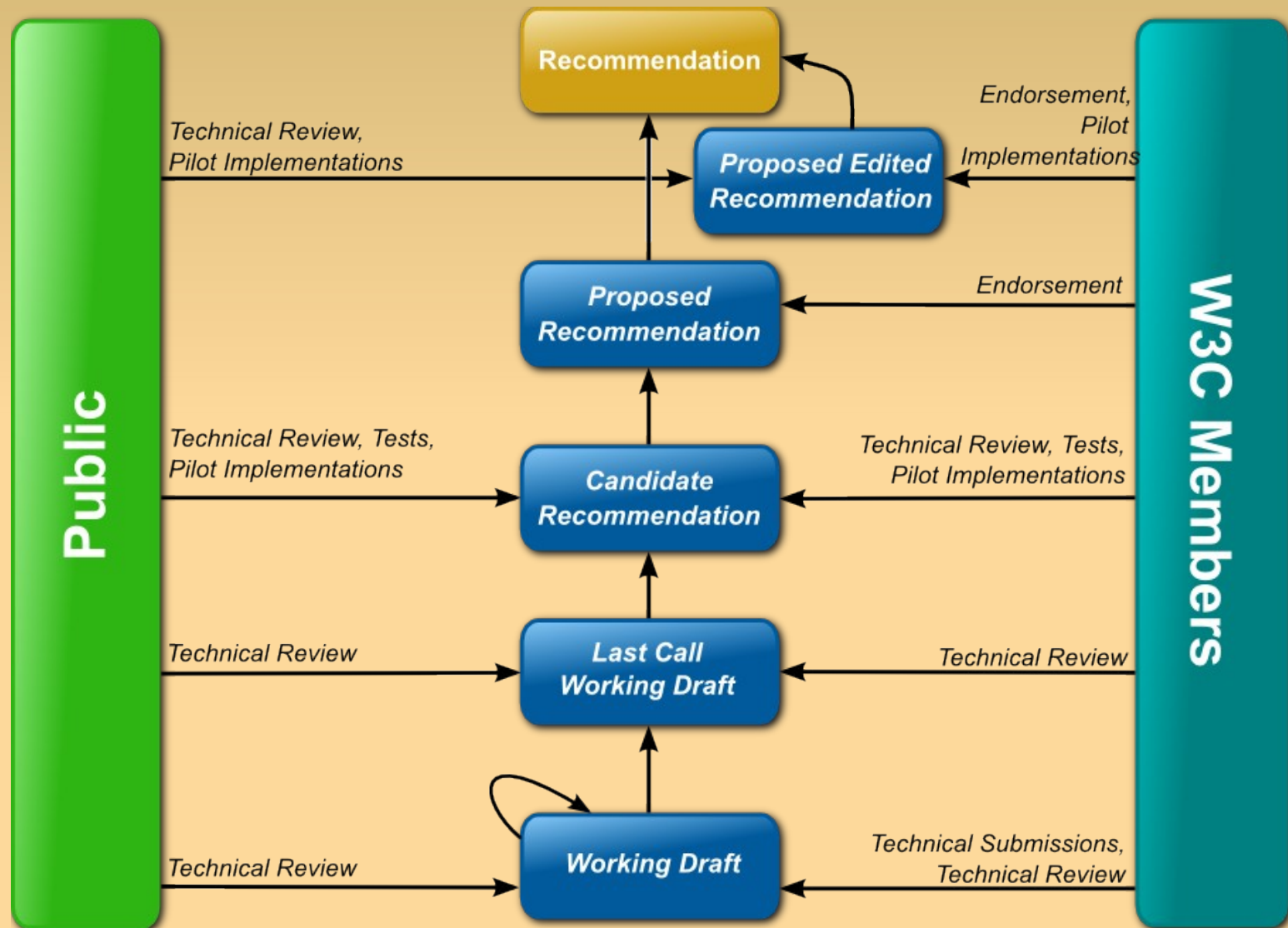
W3C's groups
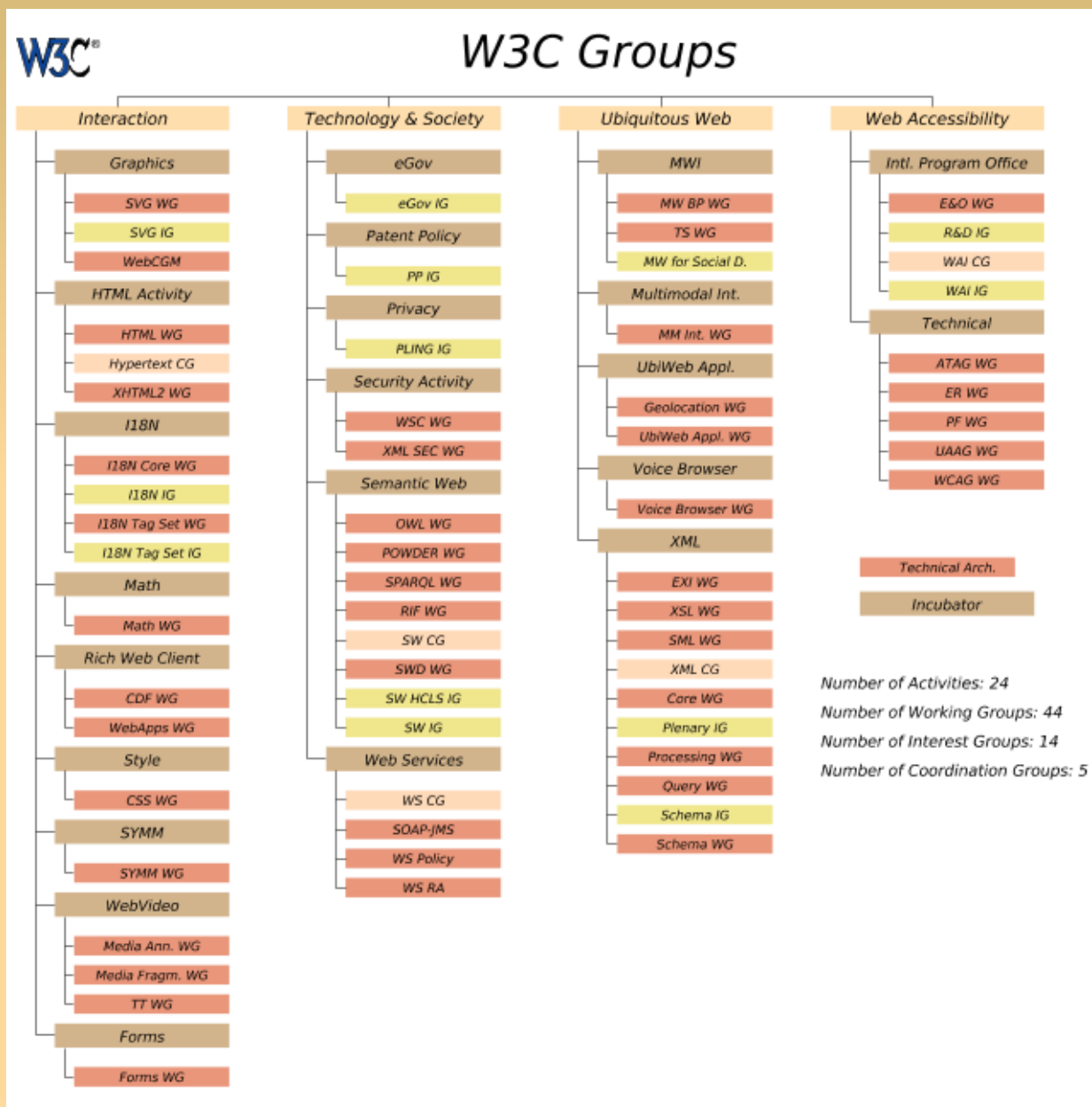
W3C Staff

62 people, 42 FTE, 25 directly in WGs

Working Group Life Cycle

W3C Recommendation Track

Details: W3C Patent Policy

- **Goal**: Produce Specs implementable on Royalty-Free basis and allow technical work with minimal interruption

- **Method**:

  - W3C Royalty-Free licensing definition: available to all, all Essential Claims 'owned or controlled', field of use limitation, reciprocity, no fees, defensive suspension.

  - W3C RF obligations for Working Group participants:

    - license Essential Claims they hold on an W3C RF basis

    - Exclusions opportunities early in the development of a specification

  - Disclosure rules for non participants, based on actual knowledge

  - Exception Handling: Patent Advisory Group (PAG)

W3C WORLD WIDE WEB
consortium

In short...

- W3C has a solid track record for developing open standard Web technologies

- Work on standards involves lots of challenging issues related to the future evolution of the Web

- The process is transparent and partially open

No Silver Bullet

- Good Design and Global Consensus take work.

Semantic Web / Linked Data

- A W3C technology stack and community, emerging from early years of the Web, gradually building momentum.

- May have some useful ideas for NoSQL

- Can probably learn a lot from NoSQL, too

W3C WORLD WIDE WEB
c o n s o r t i u m
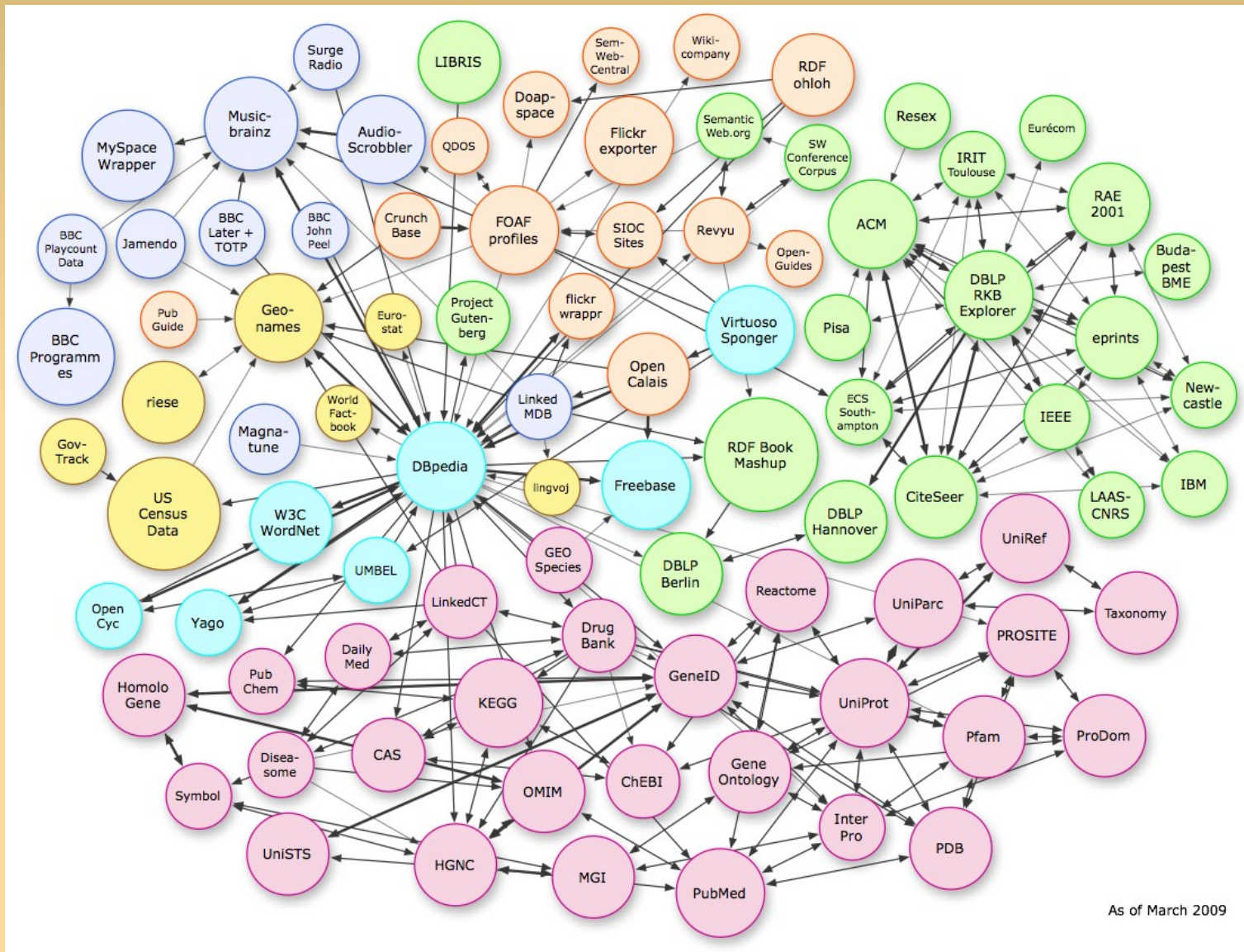
RDF Triples

- A single, standard data model
  - (Subject, Property, Value) triples
  - Often considered as a graph
  - Uses URLs as node and arc/property identifiers
  - Uses XML datatypes for literal data
  - Serialize in XML, plain text, HTML attributes, JSON
  - Libraries in C, C++, Java, Javascript, Perl, PHP, Python, Ruby, LISP, Prolog, Scala
  - Used in Adobe XMP, Oracle, Drupal, ...

Linked Data

RDF designed for **decentralized** use

- Billions of web pages, each containing a little data
- Some servers with billions of triples
- Conceptually <u>one</u> graph (linked by URIs)
- Alas, you can't query the whole thing (yet?)

Linked Open Data Cloud

As of March 2009

SPARQL

- Standard Query Language for RDF

- Triples or Quads (like Document Database)

- Roughly SQL style

```
SELECT ?homepage
FROM <http://www.w3.org/People/Berners-Lee/card>
WHERE {    card:i   foaf:knows  ?known .

    ?known   foaf:homepage  ?homepage .}
```

- Access via HTTP defined

- 24 engines listed on W3C wiki

- SPARQL 1.1 (with update! transactions?) coming in 6-12 months

Inference

- Sometimes a set of facts implies other facts

- If you know:

    - Every ipod nano has at least 1G of memory

    - My mp3 player is an ipod nano

- You can infer:

    - My mp3 player has at least 1G of memory

- Lots of work in RDF standardizing around this

- ...So it's clear which inferences are correct

- ...And sophisticated algorithms available

Things for Someone to Try

- Build SPARQL interfaces to some NoSQL systems.   (Are internal changes needed?)

- Build RDF Node-Centric interfaces to some NoSQL systems

- See how RDF-Based systems compare for various NoSQL applications (eg AllegroGraph, 4store, Virtuoso)

W3C® WORLD WIDE WEB
consortium

Conclusion

- Three Parts:

  1. Does NoSQL Need More Standards?

     Too soon to know. Some (RDF) might be helpful.

  2. How Standards are made at W3C

     Cool Tech+Enthusiasm (+time+$$) = Global Standards

  3. NoSQL, meet some RDF Standards

     → Want Decentralization?   Want Inference?

     → Collaborate on Scaling, Ease-of-Use?

W3C® WORLD WIDE WEB
consortium

More Information

- Me:
  - Sandro Hawke, sandro@w3.org
  - Twitter: sandhawke
  - Blog: **decentralyze.com**
- This Talk
  - http://www.w3.org/2010/Talks/0311-nosql
- W3C: http://www.w3.org/Consortium
- RDF: http://www.w3.org/RDF
- Mailing List: semantic-web@w3.org