Slides on the CERN-Solid collaboration

The CERN-Solid collaboration

Presentation at the Open Search SYMposium (OSSYM) 2020/10/12

Event details

Maria Dimou – CERN-Solid collaboration manager
with contributions from CERN web application developers
and Jan Schill - MSc student in the CERN-Solid code investigation
project.

World Wide Web - What happened

- The Web was invented at CERN by Sir Tim Berners-Lee (TimBL) in 1989.
- He defined it as a **free, open, networked Internet application**.
- The Web produced an uprecedented change to human civilisation.
- 30 years later, the original purpose of the Web "access to knowledge, free for all and respecting each one" is being violated.
- TimBL proposed *Solid*, the open source platform aims to give people control over their data.

The first ten years

- TimBL went to MIT to create the World Wide Web Consortium (W3C) in 1994.
- In the CERN Web Office, till 1999 we:
    - ran TimBL's httpd - followed by the Apache web server with virtual hosting, on Unix platforms.
    - negotiated a free-of-charge Netscape browser support contract.
    - deployed *pinaweb* (Personal Intelligent Newspaper Agent) (a web profile by CERN student Heidi Schuster).
    - investigated Web-based calendars
    - recommended HyperNews for collaborative work
    - and more...

Web Search at CERN

- CERN student Darius Kogut wrote *Torch*, a search engine parsing natural english language in 1998.
- This development was an intellectual satisfaction; relationship with other disciplines, the understanding of rich human language by the search engine.
- As the Web was in exponential growth, we couldn't go far with in-house development.
- So we evaluated Lycos and Altavista... they left a lot to desire.
- Finally we signed a negligeable-charge contract with *Infoseek*, then *Inktomi*. Slides from 1999

Web Search in general

- Search was "innocent" at the time;
  - companies didn't make money out of offering, withholding, manipulating information on the web.
  - Google didn't exist yet.
  - The search results were irrelevant or incomplete, still they were what existed and not what the engine would like to show the user.
  - Surveillance and intrusion were not yet terms we were conscious of.

CERN & Web standardisation since the year 2000

- Computing focus at CERN was turned to the huge amount of data produced at the LHC.
- Proposals for a CERN-W3C collaboration remained without answer for 25 years. The suggestions were:
  - to combine use of the https protocol for physics' data transfer and remote access to storage with W3C standardisation work. Details.
  - to contribute design concepts in CERN applications in areas like the Data Catalogue Vocabulary, cross-service inter-operability and Authentication/Authorisation rules and restrictions. Details.

Solid - What is

- TimBL announced the Solid project (Social Linked Data) in 2016, aiming to give people control over their data. His summary:
- This is an open source platform, adding standards never put into the original web spec, including:
  - Global single sign-on,
  - Universal access control
  - A universal data API so that any app can store data in any storage place.
- Socially, Solid is a movement towards a world in which users are in control, and empowered by large amounts of data, private, shared, and public.

CERN-Solid collaboration – born this year 2020.

CERN packages relevant to Solid spec's for evaluation:

- The CERN *push notifications*, unilateral, via subscription and archived.
- *Indico*, an event management open source platform, with 20 years of operational status.
- *CS3MESH*, a pan-European cross-institution mesh that will offer data sharing/co-editing facilities, relying on the federation of different sites by using well-known APIs.
- *InvenioRDM*, a Research Data Management, open source platform for persistent registration of research papers and data.
- *The new CERN Authentication* project.
  Web pages of the above in the "References" at the end.

Activities now

MSc student Jan Schill from https://itu.dk started working on this project. Goals include:

1. understanding which Solid specifications are ready and clear.
2. evaluating the first Solid implementations.
3. making a recommendation to the CERN open source applications on Solid adoption (or not).
4. exploring Indico, to test the Solid principles, by:
   - modifying-ala-Solid the *Indico registration form* module, so that registration data belong to the user and not to Indico.
   - enriching *Indico meeting* pages with Solid-based content, such as comments.

   Indico-related suggestions by TimBL and Pedro Ferreira.

Search-related work in Solid

- In Solid, data is stored in Pods.
- Current and planned work in Solid includes:

1. The 'Search UI' item in TimBL's roadmap (next slide) - client-side interface, for searching in one's own pod.
2. Server-side pod-wide search functionality by Fred Gibson for the TrinPod implementation of a Solid server.
3. Solid-user-search for searching a person (like in a phonebook)
4. hashtag-search, for content that is related to a given hashtag.

For design details and development status contact Michiel de Jong.

Solid (evolving) roadmap

References I

[1] The original Web proposal https://www.w3.org/History/1989/proposal.html

[2] The CERN-Solid Indico category https://indico.cern.ch/category/11962/

[3] The Solid project web site https://solidproject.org

[4] The CERN Web Office (most data missing today) https://weboffice.web.cern.ch/WebOffice/

[5] The CERN Torch search engine http://cern.ch/dimou/SApaper.html#torch

[6] CERN-W3C 2014 proposal https://cern.ch/dimou/personal/CERN-W3C_Collaboration.pdf

[7] CERN-W3C 2017 proposal https://cern.ch/dimou/personal/CERN-W3C_Collaboration_2017_proposal.pdf

References II

[8] Push notifications archaeology - proposal in 2003 https://cern.ch/dimou/it-us/zephyr.shtml

[9] Push notification proposal in 2020 https://codimd.web.cern.ch/p/ry5_j4r2U#/

[10] Linked Data Notifications: https://www.w3.org/TR/ldn/

[11] The WebSocket Protocol: https://tools.ietf.org/html/rfc6455

[12] Indico https://getindico.io/

[13] The Road to the new CERN Identification https://auth.docs.cern.ch/whitepapers/the-road-to-new-auth/

[14] CS3 MESH https://silo2.sciencedata.dk/sites/cs3mesh4eosc/

[15] InvenioRDM https://inveniosoftware.org/