

# ”The CERN-Solid collaboration project”

*Monday 12 October 2020 15:45 (15 minutes)*

Author: Maria Dimou

by Maria Dimou –CERN-Solid collaboration manager  
current data thanks to the CERN-Solid development experts.

The Web was invented at CERN by Sir Tim Berners-Lee. He defined it as a free, open and networked medium. Ever since Sir Tim went to MIT to create the World Wide Web Consortium (W3C), CERN developed a lot of important web-based applications. Still, as an organisation, CERN stayed away from the evolution of the Web, in terms of its standards and philosophy.

The management consideration was that the laboratory has to operate the LEP accelerator, prepare and build its successor, the LHC, basically to do physics. This was something also said to Sir Tim when he was programming the Delphi experiment RPCs, while the design and need of the web was clear in his mind.

In the first years of the web, when Mosaic and Netscape were still bleeding edge of web technology, in-house web development was still taking place in the, then called, CERN Web Office.

Student Heidi Schuster developed pinaweb (Personal Intelligent Newspaper Agent), a programme written in Java, guessing the user’s taste of web pages visited, creating profile per user and proposing most recent appearances on the web on the matters that interest the user subscribing to the pinaweb service. At that time manipulating via the web didn’t exist, so we found this a very clever and convenient application. Surveillance and intrusion were not yet terms we were conscious of.

Student Darius Kogut wrote Torch, a search engine understanding natural english language rather than keywords linked only with AND and OR operators. The development of this application was giving us intellectual satisfaction, as we were feeling that we were getting to grips with other disciplines, the understanding of rich human language by the search engine.

In the end, we did purchase the search engine Infoseek, later called Inktomi. They made us a good price offer, which we refused with the argument: “Your business would not exist, had the web not have been invented at CERN”. It worked. The price was symbolic.

The above projects were approved because our evaluation of Lykos, Altavista and the like was leaving a lot to desire. Also because the time was such that companies had not yet made money out of offering, withholding, manipulating information on the web. Google didn’t exist yet. The search results one was getting were probably irrelevant or incomplete, still they were what existed and not what the engine would like to show the users according to its estimation of what is appropriate/relevant/desirable for them.

By 2001 all these creative activities ended. Commercial solutions were adopted for the web matters of the lab. It is true, the CERN IT Department has to support computing applications and ensure smooth operation, and for this it is appreciated by the lab.

The years coming up to the LHC were critical because the network and storage needs were unprecedented, so we were not sure the technology would make the quantum leap before we actually needed it. Luckily it did. Today’s CERN experimental data are massive, still Google and facebook stole the first position and store even more.

Innovative and pioneer work started to be mostly expected to come from the physics arena. Technology transfer, especially usable by medical applications had first priority and most attention.

Still, there were several possibilities to link our computing developments with W3C standardisation work.

For example, in the area of the Worldwide LHC Computing Grid (WLCG) project and operations, where the use of the https protocol for data transfer and remote access to storage related naturally with the work done by the W3C Working Groups.

There are also design concepts in CERN applications, which can contribute good ideas in areas like the Data Catalogue Vocabulary, cross-service inter-operability and Authentication/Authorisation rules and restrictions.

Those CERN-W3C proposals remained without answer.

Then, Sir Tim Berners-Lee announced the Solid project in 2019. Private political or financial interests drove the web away from the principles he invented it for, namely universal, educational and free access to information. In Solid, standards will follow the principle of loose connection between Identity, Data and Applications. This will give the user control of his/her data.

At this point the climat was ripe for the CERN-Solid collaboration to be born this year.

The CERN development areas, where interesting exchanges can happen, possibly leading to selective adoption of ideas fed in / taken from the Solid specification effort are:

- The CERN push notifications project, aiming at a set of official channels that users can subscribe to, in order to receive news. The notifications are unilateral and will be 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 common feature of these applications is the user Authentication needed for their workflows for restricted data.

Ideas are discussed in a dedicated CERN-Solid gitter channel. For now the exchanges focus on Web Access Control (WAC), as implemented in CERN applications. The design of Solid Access Control Lists is attractive, because it refers to users with URIs that can live anywhere on the web.

At the time of this Abstract submission (June 2020) nothing is fixed in terms of co-development, because, in the CERN case, the data and the applications belong to CERN and not to the users, so the ACLs do exist but give no freedom to the users as to own them and decide where/how their data will be stored, indexed, accessed.

Still nothing is more useful in the area of web development than collaboration, awareness of possible malicious incentives by service providers and technical, ethical, ideological preservation of the web founding principles. Its birth place has the duty to contribute in this effort.

References:

- [0] The web original proposal by Sir Tim Berners-Lee <https://www.w3.org/History/1989/proposal.html>
- [1] Solid announcement in the press <https://www.nytimes.com/2019/11/24/opinion/world-wide-web.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](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](https://cern.ch/dimou/personal/CERN-W3C_Collaboration_2017_proposal.pdf)
- [8] Push notifications 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#/](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/>

## Preferred Contribution Format

**Presenter:** MARIA DIMOU (CERN)

**Session Classification:** Plenary Session: “Open Search Ecosystems”