

RESAW

Session 1 | Digital research infrastructure and web archives |

May-June 2013

Summary of the discussion

The following is a short summary of the most important themes debated in the first web discussion session within the RESAW project (May-June 2013). No comments have been added to the points, and focus is on the issues we have to keep in mind when taking the next steps. The themes are placed in no prioritized order.

Research infrastructure (RI).....	1
Web archives and scholars	2
Analytical tools.....	2
Preservation of research tools and research data	3
Types of web archives.....	3
Legal issues.....	3
Commercial companies as partners.....	3
Cases as examples.....	4
Epistemology and critical discussion	4
Dissemination.....	4
The following steps, the seminar and beyond.....	4

Research infrastructure (RI)

It was suggested that the ESFRI definition from the 2010 roadmap could serve as a point of departure for a definition of 'research infrastructure'. The following elements were suggested as the core: associated human resources, equipment, sets of instruments, knowledge-containing resources, structured information systems (Niels Ole).

In addition, the question was raised what the criteria are for establishing a new RI and who is to decide, and how should RESAW position itself within the ESFRI RIs? (Niels Ole)

It was maintained that interaction and interworking with other research infrastructures as well as the methodological frameworks (for studying archived web resources) are also components of a RI, and that a roadmap is crucial to sustain any infrastructure (Helen).

Finally, it was noted, that there is no single model of research infrastructure (Tobias).

Web archives and scholars

The importance of the close co-operation between web archives and scholars working with archived web material was stressed, based on the argument that the research infrastructure evolves in an interplay with scholars (Niels). And it was stressed that RIs should be community-driven and not only be considered as tools and services (Tobias) and thereby being in danger of becoming a huge institutionalised service system operating on its own terms and not research-driven (Niels Ole). In relation to this final point the long term perspective was emphasized which is a matter of maintaining insights and experiences and of financing these things (Niels Ole).

It was added that RESAW should also have close collaborations with researchers and research projects working with the live web since using live web materials can bring important research issues which will also be present while studying the archived web (e.g. building a research corpus, studying big data) (Paul, Wifak).

It was noted that, on the one hand, web archives must be the focus of RESAW, but, on the other hand, involving scholars studying the live web may be beneficial since many of the analytical tools are developed to be used on the online web (Niels).

Analytical tools

The importance of having research tools as part of RESAW was stressed (Ulf-Dietrich, Niels Ole), and the necessary competences related to the use of the tools should also be present (Niels Ole).

It was added that the task of RESAW is, on the one hand, to repurpose tools developed for analysis of online web, and, on the other, to develop new tools, specifically tailored for web archival studies (Niels).

Preservation of research tools and research data

The importance of preservation of tools and of (some) research data was stressed, since the use of digital methods in most cases also leads to the production of project specific (and tool specific) corpora (Niels Ole).

Types of web archives

The possibility of broadening the scope in terms of types of web archives was mentioned. The participating web archives are all national web archives, but it would also be interesting to experiment with including other types of web archives, for instance archives with no national obligations as to preserving the cultural heritage (Niels). This could, for instance, be some US members of the IIPC or university libraries (Helen).

Legal issues

This is definitely one of the big issues. The following points were mentioned:

- a “state of the art session” on laws at the seminar (access, use, openness, etc.) was called for (Valérie)
- several issues are involved, especially copyright and privacy (Niels)
- different legal frameworks apply in different national settings (Niels)
- legal issues may prevent existing archiving institutions to archive some trans-national materials as well as to give access to them for trans-national studies (Paul)

Commercial companies as partners

It was maintained that if RESAW becomes a Horizon 2020 application it would be interesting to include a couple of commercial companies working within the field of big data and web archiving (crawling) – Google was mentioned, and there should be other companies out there. Potential commercial partners should preferably join the workshop in December to learn about the area of web archiving and internet research (Bjarne).

To this it was noted that, on the one hand, commercial partners are not needed within the ‘Integrating Activities’ which is part of the Horizon 2020 ‘Excellent Research’ section, but, on the other, the integration of commercial

partners may add yet another good argument why a Horizon 2020 application should be supported since the possibility of cross-cutting actions to the other two Horizon 2020 sections 'Industrial leadership' and 'Societal challenges' are mentioned in the EU documents. If some of the participants know of possible – and relevant – commercial partners, please let us know, and we will invite them to come to the seminar (Niels).

Cases as examples

It was suggested that it would be good to have some case studies presented by members of the project team, both to introduce them to others, but also to demonstrate examples of practice and possibilities for structuring and organising the RESAW proposal (David, Valérie, Helen, Tobias).

Epistemology and critical discussion

A session or discussion on epistemological aspects was suggested (Valérie), and the references to critical literature on digital research infrastructure was called on (Matteo).

It was mentioned that a future task for RESAW would be to clarify the historical developments and relations between Digital Humanities and digital RI (Niels).

Dissemination

Since the number of scholars using web archives is still limited, just as knowledge of the existence of web archives is now very widespread, it is important to broaden the scope by organizing european summer schools for students, summer universities, methodological guides for scholars, a forum or mailing-list, and the like within RESAW (Valérie).

The following steps, the seminar and beyond

It was asked if the organisers of the seminar in December had any intended objectives, for instance any tangible deliverables in terms of documents/agreements, a schedule towards the production of a full project proposal, etc. (Helen).

And it was maintained that one way to organise things would be to focus on the research infrastructure RESAW intends to establish, its required components, what is already in place, what are the gaps and how we get there, which perhaps will include question of sustainability (Helen). In addition to this it was noted that a “state of the art” or “status of current affairs” from interested partners would be a good thing. Although the archives within IIPC have some common knowledge sharing information about partners and what they do is still useful for forming a potential consortium for a Horizon 2020 application (Bjarne).

Regarding the long term perspective it was mentioned that Horizon 2020 is one option (this is where it started, with the consultation last Autumn aiming at H2020), AUCost-action is another (Niels).

In continuation of this it was noted that to get RESAW really flying kick-starting it with a Horizon 2020 funded project would be preferable. Afterwards it must be driven by research institutions and web archives in cooperation. Could maybe be supported (e.g. hosting of technical infrastructure like tools, service, source code, development infrastructure) somehow afterwards by some of the existing networks within Digital Preservation (e.g. Open Planets Foundation) (Bjarne).

Finally, it was noted that it was the intention of the organisers to aim at handing in a Horizon 2020 application as soon as possible in 2014 (Niels).

Niels Brügger, June 2013