Initial Response by the CLARIAH Board

To the Comments and Recommendations by the CLARIAH International Advisory Panel in their Mid Term Evaluation Report

The CLARIAH Board thanks the IAP for their assessment of the CLARIAH project so far, of the importance that they ascribe to CLARIAH and of its leading position in Europe. It also thanks the IAP for their constructive criticism and for pointing out important aspects that must be addressed to make CLARIAH-CORE an even better project and the resulting infrastructure optimally useful for the targeted users. The CLARIAH Board has formulated an initial response to the IAP's recommendations and issues of concern here, and will try to address these issues in the second half of the project. We repeat each of the IAP's recommendations here, followed by the response by the CLARIAH Board

1. Integration should be the next priority, and there should be more concrete plans on how to carry out integration of the three Work Packages. Integration of additional humanities fields, however, seems to us to be a lower priority.
   - The Board agrees that integration of the 3 disciplines is important. Since CLARIAH brings together three fields that were almost completely separated before the start of the project both scientifically, methodologically and in terms of infrastructure, this is a huge task that will take considerable time and effort. The board is already working on integration through cross-disciplinary projects covering 3 disciplines (in particular, linguistics-social economic history and linguistics-media studies); the Athena project, which was not prominently discussed at the meeting also offers a lot of potential for the integration of the three disciplines. Common interests have also been identified and explored at the various workshops that have been organized, such as those on Linked Open Data and Interoperability. The board will investigate other opportunities for further integration.
2. We suggest that you try to identify a single topic that can work well for integration of the three fields. While the demonstration of successful integration of this single topic would be too narrow a base for the entire project, we think it would be a good test case.
   - A single topic might indeed be very helpful for further integration. We have in the past discussed several such topics at different occasions:
     i. The Athena dataset on historical information on the environment (see previous point).

ii.   At the presentation of CLARIAH (CORE) for the granting committee, we discussed migration. Migration and its effects on society, economy, distribution of wealth, and language, as well as the representation of these matters in the media is an excellent example. More generally, the Horizon2020 challenge to strive for an inclusive society (i.e. a society for *all* its citizens)[1] is a challenge indeed, with the society not only confronting issues such as increasing migration and increased interest for (perceived or real) national identities as mentioned above, but also multilingualism, problems of language education, increased stigmatisation and polarisation, terrorism, disruptive economies made possible by the digital information society, for example an ongoing redistribution between information exchange through traditional media (newspapers, journals, Radio, TV) v. information exchange via new social media (Twitter, Facebook YouTube, etc.), and a highly skewed distribution of wealth and income.

iii.   An application for an NWO Zwaartekracht (gravity) grant was explored around the theme of radicalisation, where the existence of the CLARIAH data set was one of the foundational principles for the cooperation. Eventually this was not pursued in 2016 because of time constraints. Zwaartekracht proposals are very broad, but also very ambitious. We will re-consider this option if and when the Zwaartekracht programme is re-opened for applications in the future.

iv.   Meertens Institute and Huygens ING are considering a research project on national identity, which is a very suitable theme for a CLARIAH-wide effort.

As these examples underline, such a venture would have to rely on a separate, external source of funding (1, 3, 4). We gave examples of projects that CLARIAH will make possible when we applied for our present funding (2). However these funds have been granted to build an infrastructure, and we cannot relocate to do the practical research. Parts of the infrastructure can point to specific research topics (1), but they are not research activities in themselves. If we apply for new funding for a specific research project, the right coalition of applicants will normally be determined by the topic, and ne less broad and more focussed than the CLARIAH community as a whole (1, 4). They will want to select data and tools that serve their research interests, not showcase CLARIAH.

This leads to the conclusion that to showcase CLARIAH, we will have to apply for a specific project that is suitable for this purpose, and form a group of applicants where this is one of the agreed goals from the stat (like 3). As mentioned (in 3, above), there are plans to do so. This, however depends on the specific Zwaartekracht call for proposals being opened. We will try to develop additional plans for other calls.

3. WP2 seems to be behind the other WPs in the degree of its progress. Readers of CLARIAH reports and users of CLARIAH interfaces will see the gaps, for instance in incomplete standardization of metadata.
   ○ The impression of the IAP that WP2 is more behind than other WPs is not correct in our view. Though WP2 so far may appear to have spent relatively little money, this is in part due to delayed invoicing. WP2's activities are reasonably on schedule, and the situation is not significantly different than e.g. in WP5.
4. As is common in academic projects, the CLARIAH PIs must wear many hats, as they carry responsibility for conceptualization, administration, reporting, and so forth. At later stages of the project, including integration as well as completing the infrastructure, this multi-tasking will be compounded. The problem won't go away, but conscious attention to it may help save time for attention to the highest priorities.
   ○ This recommendation will be attended to by the CLARIAH Board.
5. We applaud the strategy of CLARIAH, attempting to merge at the Netherlands national level the approaches of CLARIN and DARIAH at the European level. In a real sense CLARIAH is a test case for all of Europe. Especially since members of CLARIAH are on boards overlapping with CLARIN and DARIAH, we encourage the opening of discussion with members of those two bodies even before the completion of the initial CLARIAH project, to see if they have helpful suggestions and to enable them to envision the steps that they too will encounter.
   ○ We are happy with the observation by the IAP, which indeed we observed ourselves as well, especially very prominently at the DARIAH Annual Event in Berlin (25-26 April 2017). We already started investigating the organisation of a joint event with other countries to discuss opportunities and challenges for integrating CLARIN and DARIAH nationally.
6. We note two types of appointments for which there are bottlenecks or shortages. The emergence of academic technology specialists as mediators between researchers and technicians is promising in general, though few have been trained in this area. But the real problem for CLARIAH seems to be the shortage of engineers. Perhaps extra efforts can be made to make their jobs rewarding, integrating them as fully as possible into each WP.
   ○ The matter of getting and keeping staff, especially engineers and ICT researchers has our continuous attention but there is very little we can do financially or in terms of permanency of positions. However we must ensure that the working conditions are optimal and create an intellectually challenging working environment, which may attract engineers and ICT researchers to opt for CLARIAH rather than for positions in commercial companies. There seems to be a class of software engineers that considers cooperation with (humanities) researchers as a great asset. Close cooperation between researchers and engineers, as is now happening within the research pilots, may very well help to keep engineers interested in these tasks. Outsourcing some work to commercial companies was suggested as an option, which it may be (and is actually done) in some cases, but because of continuity, sustainability and financial restrictions it

cannot be considered the ideal solution for all cases, in particular concerning essential components of the architecture.

7. The interfaces for the tools seem to be neglected, though less so for audio-visual than for linguistics and socio-economic history. Can it be that technicians create the interface and researchers are involved only after it is completed? Clarity of access and access to the granularity of research issues should be a top priority in each WP, through an interactive design process.

   - Concerning the interfaces of tools: we fully agree that having good user interfaces is very important for the successful application of tools and services by the targeted users, humanities researchers. The demo session may in some respects have been less representative for the work done in CLARIAH. For WP3 we indeed illustrated aspects of the infrastructure where no good user interfaces are available yet, in part because this functionality will be hidden behind other applications, and in part because we did not want to demonstrate tools and services that do have a dedicated user interface again that we already demonstrated at the IAP meeting in January 2016. In many of the applications shown there actually multiple different but interrelated interfaces are offered, for different types of users and for different types of query complexity. In the WP3 Search projects that have recently started at the INT software- and interface development will be carried out in close cooperation with targeted users, in particular the PhD students of the Dutch national research school in linguistics (LOT). In WP5, there is full attention for good user interfaces, including pre-defined but modifiable recipes for novice users, but the focus so far has been on integrating the backend functionality of existing applications into a single media suite; the user interfaces will be further developed once the tools have been tested. In WP2 and WP4 most attention so far has gone to creating back-end functionality and curating data, which is a prerequisite for applications with dedicated user interfaces to make use of these data and functionality. In addition, its core services are also primarily aimed at engineers working in the humanities - instead of scholars and researchers. Plans for an overall interface or interfaces, comparable to the media suite, are under consideration, also as a separate, new task in CLARIAH-PLUS.

8. The interim report and draft proposal both read from the top down. Perhaps some lower-level or more granular examples can be given early in the report, to give readers a sense of the multiple levels of the project and to enable them to remember engaging examples.

   - This could be seen in connection with points 1 and 2. If we have a single topic for integration, this concrete topic might serve as a lower level example to provide an integrated description of the work that has been carried out in CLARIAH. We will devote attention to the top-down impression given by our texts.

9. I would find it helpful if each portion of the report gave a sense of how many people are working in each area. Each WP introduces itself as "we," but this is vague in number and in the specializations of those in the group. It would be helpful to know the proportion of

researchers and technicians in various project groups, to get a sense of the agency of project staff.
- ○ We will inventory this information and report on it.
10. Connections among the WPs can perhaps be emphasized on the CLARIAH website and even within the reports.
    - ○ We will create more publicity on integration show cases
11. Workshops, as emphasized by WP4 and WP5, seem very promising. These can build recognition and support for CLARIAH, helpful in convincing funders that the project will have a public effect. Workshops could be on the main focus of each WP, though I noted how WP4 might provide an important service by finding ways to link, in workshops, historians who work with structured data and those who work with texts. In addition, such a workshop might also show the potential for inclusion of audio-visual materials from WP5. Finally, the idea of WP3 about workshops on interoperability seems very promising.
    - ○ We will step up the educational and dissemination programme, which involves not only workshops but also tutorials (e.g. at seasonal schools) as well as 'evangelisation' activities at conferences and workshops of the targeted users: linguists, social economic historians and media studies scholars.

---

[1]
https://ec.europa.eu/programmes/horizon2020/en/h2020-section/europe-changing-world-inclusive-innovative-and-reflective-societies