

**COST Action no. IS1310**

**Action Title: Reassembling the Republic of Letters, 1500-1800 A digital framework for multi-lateral collaboration on Europe's intellectual history**

**Working Group 5 Meeting**

**Venue Oxford, Country: UK, Wednesday September 2<sup>nd</sup>, 2015, 09.00 – 15.00**

**Minutes by Arno Bosse, University of Oxford**

- 1) The Chair (Thomas Stäcker) welcomed participants. Ten WG5 members, five local members (University of Oxford) and one invited external speaker (Cambridge University) were present. A representative from Europeana Professional accompanied part of the discussion via teleconference (see Annex for a full list of participants).
- 2) The agenda prepared by the University of Oxford was adopted (see Annex). Thomas Stäcker turned the chair of the meeting over to Howard Hotson.

**3) I. Identifying Problems: what difficulties is our new infrastructure designed to overcome?**

Consensus was quickly reached around the description of the difficulties outlined in the previously circulated WG discussion paper (see Annex). The key challenges identified in the paper were: i) Different modes of sharing data, ii) Different kinds of data shared, iii) Decentralized data curation, iv) Centralized interrogation of distributed data. It was decided to move on to the second item in the agenda.

**4) II. Technical means: what kind of infrastructure can achieve these aims?**

a) What are the chief challenges to structuring and sharing correspondence records as Linked Open Data (LOD)?

Licensing and copyright issues were raised ("What you can do with the licensing determines what you can do with the data"). Does EMLO has significant publishing restrictions? No, since it uses top level metadata and links to transcripts or page images. But it will be a problem for working with transcripts and images. Some data providers will be restricted in their ability to share metadata due to existing copyright restrictions.

Would CERL ("Council of European Research Libraries") be able to ingest data from a revamped EMLO? Enriched metadata can't be sent back to the original institution if they are still using MARC etc. Enrichments could be kept and displayed in a parallel system? But a library still needs to do the work to link to resources.

**Task:** Ivan Boserup will investigate if CERL could share and ingest LOD metadata.

The distinction was raised between sharing a copy of the data and sharing an index created from the data. In the UK it is possible to make a text available for machine reading (indexing) but not human reading. A national library as a hub can resolve these kinds of copyright problems far better than a central instance.

b) If we envision the infrastructure as a network of cloned nodes then how many nodes, in what sorts of institutions, and how much cloned functionality in each?

The idea of providing a turnkey solution for a LOD node (“EMLO-in-a-box”) was raised. For a small institution, this would consist of a VM environment (which could be self-hosted to managed by an external provider) with all the necessary software, documentation, protocols and standards already pre-configured to become an active LOD node. For a large institution, the documentation, protocols and standards included in the VM would still be useful as a reference against which to test their own implementation.

**Task:** Neil Jeffries will prepare a short white-paper on EMLO-in-a-box.

c) Quality control and accreditation. How do we build quality control into the system?

The seven star model was discussed which builds on Tim Bernard-Lee’s 5-star model for the semantic web. An additional star is given for documentation of content and another for documentation of process / quality and provenance. However, the consensus was that quality control and provenance hasn’t been solved although inroads have been made. A good way to capture, track and publish credit for enrichments will need to be provided.

This led into a discussion on authorization and authentication. Individuals move from institution to institution and from reputation to reputation. How do we work around people who aren’t members of these institutions? We also need identifiers for people who are no longer alive, and from contributors we can’t identify. A suggested solution was to decentralize authorization/permissions but centralize authentication/identity.

**Task:** Neil Jeffries will report back on his discussions with Project THOR (<http://project-thor.eu>) and ORCID (<http://orcid.org>).

d) How do we address the problem of long-term sustainability and preservation?

Is there a danger that we be used as a preservation & sustainability dump for some projects? CLARIN has sustainability integrated into the infrastructure. An aggregator like Europeana can also be a part of a broader sustainability solution. The project could share all of its data as dark data for preservation purposes.

e) Will this solution meet the needs of major repositories? What is the added value we provide to scholars and institutional data contributors?

Successful reconciliation is a major benefit and can be passed on as enrichment. Given an ability to track provenance, individuals could demonstrate their contributions to their committee. What can we do to assign a unique ID to every letter in Europe?

**Task:** Thomas Stacker will investigate the use of identifiers for work, items etc. roughly following the FRBR model.

**Task:** Oxford will prepare a paper on the ‘business case’: How do libraries & other data contributors benefit?

**Task:** Thomas Wallnig will provide a series of scholarly/research case studies as an outcome of the upcoming Vienna COST Action meeting.

f) Letter texts, analytical, transcription, and annotation tools. Can we agree on basic priorities and specifications? What about browsing, querying and visualization tools?

It was agreed that these questions were too large for this meeting and better addressed by the appropriate Working Groups (WG3 and WG6).

**Task:** Charles van den Heuvel will discuss visualization and user-interfaces with Paolo Ciuccarelli and prepare a synopsis of current text tools (transcription, annotation, etc.).

### 5) III. Grant planning: What sources of funding are available in general, and how do we pursue them?

a) Is an Horizon 2020 “starting communities” research infrastructure (RI) grant an appropriate funding source?

There was a strong consensus that preparations should be made to apply for Horizon 2020 funding.

b) Should we expand the scope of our proposal (section VII of the discussion paper)? If so, how, and how far?

While a significant scope will be important to the H2020 proposal, there is also a danger of including too much in an e-infrastructure definition. The UK government, for example, thinks it's just supercomputers. We're not, by some measures, empirically big (e.g. in terms of data volume) but our metadata and analysis needs are nonetheless complex.

c) What forms of involvement do we anticipate for various members of the COST network? And how do we determine which form of involvement is appropriate in each case?

We need a roving ambassador to help find and speak to the institutions that want to set themselves up as hubs. Within the COST membership, we need know on what basis we should talk with our national contact points.

**Task:** Oxford will prepare a 1-3pp “talking points” document for respective national contact points explaining the COST Action and the proposed infrastructure ‘starting community’.

c) How do we address the request (possibly requirement) to incorporate industry into the proposal?

ESFRI understands that different scientific communities require different levels of engagement with the private sector. A private company could help to build parts of the infrastructure. Ontotext has collaborated a lot with different Europeana projects and/or participated in EU projects. SYSTAP (‘Blazegraph’ graph database) collaborates with Wikidata, OpenLink works with DBpedia and on various EU projects.

### 6) Closing

A Google Doc was shared with the participants with rough notes on the discussion.

Thomas Stäcker and Howard Hotson thanked all participants for a fruitful, stimulating and productive meeting.