Data Sharing Agreement

As a research infrastructure project funded by the European Commission and a formational contributor to the ESFRI roadmap through its parent infrastructure, DARIAH, CENDARI's primary concerns are to facilitate the access to and interrogation of the primary data required for humanistic research. CENDARI is not intended as an access portal for the general public (though members of the public may find and use it), nor as a primary delivery mechanism for a broad range of cultural content (though a lot of content will be available through it). Instead, CENDARI's primary role is to become a powerful workbench for the deep investigation of humanistic questions, capturing the tacit knowledge of the scholar-user and the archivist and enhancing those perspectives through its data curating, mining, query and visualization tools.

As such, the project places a very high value on *attribution* of the data it delivers, creating traceability for all elements in the infrastructure back to a single source. We strongly believe this is the only way to create a scholarly digital ecosystem with equivalent conditions to those of traditional research environments for the results to be transparent and reproduceable, the most meaningful quality standard for any scholarly work.

A single project cannot mandate usage and attribution patterns for all data and all institutions. What follows here, however, are the principles by which CENDARI will determine appropriate standards for sharing and exposing content clearly within its own technical control, and what the project's baseline recommendation to its scholarly and archival collaborators and contributors in this respect will be. As such, CENDARI hopes to set a precedent for sustainable good practice within digital humanities research and research infrastructure.

This good practice will be based on a commitment to the widest possible usage of the CC-BY standard. CC-BY (Creative Commons Attribution 2.0 Generic) allows the user/researcher is free to share (to copy, distribute and transmit the work), to remix (to adapt the work) and to make commercial use of the work under the condition that the work is attributed in the manner specified by the author or licensor. See: http://creativecommons.org/licenses/by/3.0/

1. Data or documentation created by CENDARI for CENDARI: this data will be available automatically for use under a Creative Commons CC-BY license, with reference back to CENDARI itself or the contributing partner.

2. Publicly visible data created by scholar-users within CENDARI and disseminated through its platform: this data will also be available under a Creative Commons CC-BY license, with reference back to the author. User-data within CENDARI will only be made public at the user/creator's request and this standard need not be applied in the case of data created using CENDARI tools or content, but hosted and

disseminated independently from it, or to data created and hosted within CENDARI but only held for private visibility.

3. To maintain this standard, archival metadata and digital content available through CENDARI will also need to be available as CC-BY, attributable back to the partner archive. We realize that this may be a complicated standard for some archives to apply to some data, and will work with our partners to ensure maximal inclusion of their data that does not create conflicts with any other preexisting copyright agreements. As CENDARI will not normally be hosting full data-sets from participating archives, but only a registry, lower-quality facsimiles (eg image thumbnails), metadata, links etc., CENDARI's services to partner archives will be limited to accurately reflecting holdings, and providing access to them in seamless combination with other, cognate collections. Every effort will be made to respond quickly to any problems in the CENDARI register of partner data, to protect the rights of the content owners, including but not limited to maintaining data security, and to ensure that data provenance is clearly reflected within CENDARI. Any required transformations to archive data will be specifically agreed with the archive in question, and the results of any enrichments will be shared at no cost with the partner archives.