Bard Graduate Center
The Distributed Text:
An Annotated Digital Edition of Franz Boas's Pioneering Ethnography

Data Management Plan

The two categories of data generated in the course of this project will be managed, for the duration of project development, via the project wiki as explained in the Narrative and below, and disseminated on the final digital edition website as described in the Narrative. The first category consists of digitized materials related to the original text and intended for use on the website as primary or interpretive content. The originals of these materials were created in diverse media and include manuscripts and typescripts in English, Kwak'wala, and German; historical and recent photographs; newspaper clippings; sound recordings; video clips from various sources, including interviews with contemporary knowledge-bearers; and published essays, narratives, songs, and images. The project team has already amassed a significant quantity of information about the data to be included in this category, in the form of institutional catalogue information, photocopies, photographs, and digital files. The Start-Up phase will result in as complete an inventory as possible of these materials in preparation for their digitization.

The second category of data relates to open-source software created in the course of preparing the website: fonts; search algorithms; and software created to display, digitize or transliterate Kwak'wala. Upon completion of the project, fonts and software will be made freely available from the website and from institutional project partners. Information gathered in the course of developing the linguistic software, and in developing website design and functionality, will be reported in the white paper submitted at the end of the Start-Up period.

Early in the Start-Up phase, Berman and Glass will use structured pages on the project wiki (hosted on the Wikidot.com service by the BGC as an initial database. This will allow the team to immediately begin this inventory and upload existing and emerging digital files. The database may be shifted to a program such as Filemaker Pro for more granular and searchable data organization. The wiki will also serve as the foundation for website structure and to prepare for larger-scale digitization. The wiki/database will track metadata relating to the objects' origins (creator, date of creation, type of media, holding institution, language), the project's tasks (obtaining reproduction rights, digitization status and format, translation from Kwak'wala or German, transcription from audio or video sources), and the structure of the information (page reference in the original 1897 text or cross-references to other materials in the text, the database, or external content such as published academic literature). As possible users of this database include the holding institutions, the metadata scheme will be designed in consultation with them, perhaps using an interoperable, cross-disciplinary metadata standard like Dublin Core.

The Wikidot service provides robust networked backup. In addition, periodic archival copies of the wiki (and if necessary, separate database) will be made and stored as files or on optical media. According to the standard practice of the BGC Digital Media Lab, alternate copies will be made on a local computer on a weekly basis to ensure the database survives possible server failure.

Some data have potential issues of confidentiality and cultural sensitivity, such as use of Kwakwaka'wakw proprietary knowledge and disclosure of genealogical or other personal information. The project's Kwakwaka'wakw partners will help develop guidelines for such potentially sensitive areas. Because the wiki contents will be accessible to project advisors and digital technology consultants, these guidelines might include restricting access to certain portions of the database. Such decisions will also apply to the final public website, and the wiki building phase provides the opportunity to address them.

Outside of these constraints, the goal of the project is to make its data accessible to all. Ultimately, a version of the database will be incorporated into the public website. The full version will be preserved and accessible via the project wiki on Wikidot.com servers or as a Filemaker Pro file stored on the BGC servers and possibly transferred to the U'mista Cultural Centre and other partner institutions with confidentiality agreements.

IP rights for the primary project data (archival materials) will be determined on a case-by-case basis with holding institutions prior to digitization or incorporation on the wiki. Many materials (such as Boas and Hunt's published texts) are in the public domain. IP rights for the final digital edition will be determined based on the identity of the institution that will host and maintain the website; this should be clear by the end of the Start-Up period.