Connexions module: m14159

THE NATURE OF HUMANITIES AND SOCIAL SCIENCE DATA*

American Council of Learned Societies Commission on Cyberinfrastructure for the Huma

This work is produced by The Connexions Project and licensed under the Creative Commons Attribution License †

1 The Nature of Humanities and Social Science Data

Digitizing the products of human culture and society poses intrinsic problems of complexity and scale. The complexity of the record of human cultures—a record that is multilingual, historically specific, geographically dispersed, and often highly ambiguous in meaning—makes digitization difficult and expensive. Moreover, a critical mass of information is often necessary for understanding both the context and the specifics of an artifact or event, and this may include large collections of multimedia content: images, text, moving images, audio. Humanities scholars are often concerned with how meaning is created, communicated, manipulated, and perceived. Recent trends in scholarship have broadened the sense of what falls within a given academic discipline: for example, scholars who in the past might have worked only with texts now turn to architecture and urban planning, art, music, video games, film and television, fashion illustrations, billboards, dance videos, graffiti, and blogs.

The archive of the University of Southern California's USC Shoah Foundation Institute for Visual History and Education ¹ is a good example of the value of critical mass or functional completeness. The tale of what happened to one or two families, in one or two villages, in one or two countries, during the Holocaust is worth recording and disseminating. But we can gain far more knowledge from the record of some fifty-two thousand testimonies. In history, art history, classics, or any other scholarly enterprise that benefits from a comprehensive comparative approach, quantity can become quality.

The problems of digitizing cultural documents are multiplied when these documents have many audiences. Within the social sciences and humanities, there can be numerous subject specialists who want access to the same sources for different reasons. For example, the Roman de la Rose Project, a stunning digital collection of the major illuminated manuscripts of the Roman de la Rose, a popular medieval literary work, ²is used by literary scholars, art historians, linguists, social historians, and preservation specialists, each of whom has a different disciplinary perspective and vocabulary. Students and the general public often use such documents as well, and since those audiences want further contextualization, the data or evidence itself needs to carry, within itself, more self-description and more cues about the context in which it belongs.

^{*}Version 1.1: Dec 11, 2006 3:39 pm -0600

[†]http://creativecommons.org/licenses/by/2.0/

¹ http://www.usc.edu/schools/college/vhi/ (<http://www.usc.edu/schools/college/vhi/>).

²Johns Hopkins University and the Pierpont Morgan Library, Roman de la Rose http://rose.mse.jhu.edu/(<http://rose.mse.jhu.edu/>).