

# Statistical Studies of Microfilm Digitization Service Output: A Survey of Vendor Capability

Ronald J. Murray; Library of Congress; Washington DC, USA

## Abstract

*The quality, fitness for purpose, and cost of digitized microfilm images and their accompanying metadata (or indeed any outsourced analog or digital product) are typically specified by contractual arrangements with vendors, who are then free to produce conforming deliverables in any manner they see fit. Given this degree of freedom accorded vendors with respect to digital image and metadata production, it becomes essential that the vendor and its library, archive, or museum clients be able to demonstrate that conformance to contracted specifications is (a.) achievable given the technology available to the vendor or in the technology marketplace, and (b.) that it has been achieved with the deliverables at hand.*

*To support a process control and quality improvement process that results in a mission-critical degree of assurance among libraries, archives, museums, and the vendors who provide microfilm digitization services, and to provide a model that interested parties in the Cultural Heritage community can study and improve upon, a series of statistical studies of the microfilm digitization process is being undertaken. The studies begin by modeling the entire microfilm digitization process as a series of coordinated steps – with newspaper issues and reels of microfilm as inputs and outputs – that result in the production of digital images of specified quality. The inputs or outputs of selected steps in the microfilm digitization process are then chosen for their importance to the process as well as their being amenable to qualitative or quantitative measurement.*

*The final result of the processing steps – a digital image – is intended to serve purposes that will be dependent on characteristics of the digital image. Casual or close reading, examination of newspaper typeface characteristics and paper textures or color, and the location of textual passages using full-text indexing can be found to require images of differing spatial and tonal resolution. In order to predict during the production process the fitness of a digital image of a microfilm frame for the purposes intended by a Cultural Heritage institution, it will be necessary to determine whether there exists a predictable relationship between the measured or judged quality of the original materials, its microfilmed images, the digital image and the outputs of selected production steps that lead to the creation of the digitized image.*

*With improved predictability of digital imaging outcomes, opportunities therefore exist for identifying and eliminating undesirable digitization process variation – and for introducing corrective process steps for managing printed materials that are determined to be (via a human judgment or a mechanical measurement step) damaged or otherwise reduced in quality. Before this level of understanding and control of a digital imaging process can be achieved, characterization of all critical steps in the processing chain must occur, so that their effects on the quality of the resulting digitized images can be appreciated.*

*This first study involves gaining an initial indication of the imaging capabilities and range of normal variation of commonly used microfilm digitization devices and their associated imaging practices independently of the physical characteristics of any microfilm reel that will be digitized. With the capabilities and range of variation of the imaging devices known, the characteristics and range of variation of the microfilm materials themselves can then be determined. With this knowledge, three possibilities present themselves: the ability of a quality monitoring regime to detect genuine variation in the digitization process can be estimated; the benefits of improving specific aspects in the digitization process – notably improvements in imaging devices and imaging practices – can be determined; and the upper limits of quality-related and quality-limited deliverable characteristics such as OCR accuracy and judged/printed newspaper image quality can then be established.*

*This first study reports on the collection and analysis of quantitative measures of image quality made using standardized testing materials supplied to vendors who provide microfilm digitization services to the Cultural Heritage community.*

*The results of the first study indicate that additional training may be needed to establish a proper understanding (among management and production staff alike), acceptance, and execution of quality control procedures such as device characterization, the use of targets during actual production, corrective and preventive action, automatic and/or manual recordkeeping, and the like. Also indicated is the need for digitization technology manufacturers and the vendors who use their products to participate in the quality control process by collecting and providing device characterization data as well as by designing in product features that simplify the collection of quality control data.*