

## Segmenting the Government Information Corpus by Shari Laster

Government information librarians are not alone in our need for a long-term strategy to ensure the future availability of resources, but it can be a challenge to find common ground with other librarians. After all, our collections are delineated first by provenance, rather than format, content, audience, usage, or historical significance; and are set further apart by the imperative to ensure long-term access beyond the research elite. At the same time, we cannot identify a path forward in isolation: we must seek new approaches and alliances to achieve the unquestionably lofty goals that we, as a community, have set for ourselves.

A small body of research examines the relationship between a hypothetical known number of carefully maintained print copies of a digitized work and the probability of an intact print copy being available for access at a specific point in the future. Much of this research focuses on a subset of the scholarly or academic journal corpus, and relies on an infrastructure that is not applicable to the corpus of government information. This corpus includes materials under the umbrella of the forthcoming Federal Depository Library Program's National Preservation Plan, as well as other resources from the broader landscape of information products produced by governmental and quasi-governmental entities.

Rather than seeking a universal and uniform solution, I suggest we encourage and build upon a more nuanced approach. While there is both logic and value to viewing the corpus as a set of collections based on the originating governmental unit, focusing on additional characteristics can help us identify useful congruencies with other library collections. Portions of the corpus are already managed in tandem with non-government resources, including maps, posters, and even social media. We can look at these and other existing models for inspiration and motivation, and identify additional materials that will benefit from appropriately adapted versions of approaches used for other library collections.

As a thought exercise, consider the corpus of government information as if it existed in an enormous multidimensional matrix or array. Each *dimension* describes a different type of characteristic, and is composed of widely-recognized *categories* that group similar objects together. Ideally, the categories are composed such that an individual resource, identified at the instantiation level (i.e., print version, microfiche version, and so on), is primarily associated with one *characteristic* for each dimension.

For a non-exhaustive example of dimensions and categories, the following array of typical characteristics might be used to describe specific United States government information resources in tangible formats:

**Format:** book, pamphlet, map, poster, looseleaf, magazine, electronic media, videotape, floppy disk, reel-to-reel, microformat, manipulative, multiformat...

**Permanence:** hardbound, paperbound, bound with other volumes, stapled, hole-punched, ephemeral, diazo acetate, diazo polyester, silver halide polyester, legacy technology...

**Distribution:** pre-FDLP (prior to the Printing Act of 1895), early FDLP (prior to the Depository Library Act of 1962), modern FDLP: high selection rate, modern FDLP: low selection rate, non-depository publication, commercial republication, deposited with the National Archives...

**Library treatment:** annual, serial, monograph, superseded, updated (looseleaf), discarded, special collections...

**Access patterns:** used by general public, used by specialized audience, educational use, historical research, prospective research, public record, curiosity...

**Digital versions:** born-digital, access-level digitization, format-specific barrier to digital access, content-specific barrier to digital access...

**Other versions:** no other version, multiple editions, multiple formats...

In a viable array, every tangible government information resource would be most strongly associated with one characteristic in each dimension. For example, a print journal could be described as a magazine that was distributed within the FDLP, cataloged as a serial and bound with other volumes, used by a specialized audience, and possibly made available in an access-level digital version of unknown quality. Applying a multidimensional taxonomy in this manner groups materials together by the characteristics they have in common other than provenance.

In order to identify congruencies with other collections, we can focus on relevant subsets of dimensions. As an example of a subset, consider the dimensions of *format*, *permanence*, and *access patterns*. A staple-bound Agricultural Research Service bulletin from the 1950s may have been created to communicate state-of-the-art developments in food preservation, but now is additionally valuable for shedding light on the role of women in American households during this period. Following preservation-level digitization and deposit into libraries' digital archives, print objects might be managed in archival collections, ideally with waivers to lend to other libraries under controlled conditions. A paperbound Library of Congress annotated bibliography from the 1980s may be a frequent candidate for withdrawal at libraries, given the current reliance on digital tools for information discovery, but would be of interest to those conducting historical research. Shared print management strategies already in use for other materials could provide a model to scale up to the national level.

While this approach as outlined above still leaves much to be desired in terms of large-scale applicability, current challenges demand that we look for next steps that are both responsible and manageable. It is vital to keep in mind that every item in this hypothetical array has at least one shared characteristic: it is an information resource produced with taxpayer funding, and part of the enormously rich record of the activities and functions of the government. This consideration weighs more heavily on both the least glamorous and most endangered content within the corpus. Our responsibility is to proactively advocate for these materials in ways that are responsive to the surrounding environment while adequately representing the crucial role of libraries in preserving and providing access to these resources.