

UC San Diego

UC San Diego Previously Published Works

Title

The Digital-Surrogate Seal of Approval: a Consumer-oriented Standard

Permalink

<https://escholarship.org/uc/item/8cm357g3>

Journal

D-Lib Magazine, 19(3)

ISSN

1082-9873

Authors

Jacobs, James A

Jacobs, James R

Publication Date

2013-03-01

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-ShareAlike License, available at <https://creativecommons.org/licenses/by-nc-sa/3.0/>

Peer reviewed

D-Lib Magazine

March/April 2013
Volume 19, Number 3/4

The Digital-Surrogate Seal of Approval: a Consumer-oriented Standard

James A. Jacobs
University of California San Diego
jajacobs@ucsd.edu

James R. Jacobs
Stanford University
jrjacobs@stanford.edu

doi:10.1045/march2013-jacobs

Abstract

We propose the "Digital-Surrogate Seal of Approval" (DSSOA) as a simple way of describing digital objects created from printed books and other non-digital originals as surrogates for the analog original. The DSSOA denotes that a digitization accurately and completely replicates the content and presentation of the original. It can be used to express an intended goal during the planning stages of digitization and to guarantee the quality of existing digital surrogates. The DSSOA Criteria can be used to evaluate individual digital objects or entire completed collections. DSSOA is independent of production technologies and methodologies and focuses instead on the perspective of consumers – including libraries that rely on digital surrogates.

Justification of a New Standard

Digitization of existing books and other static, born-analog materials such as magazines, newspapers, and journals is increasingly valued and important for libraries and the communities they serve. This is true because digitization promises a lot: better access (faster, distributed, simultaneous), better preservation (protection of brittle originals from deterioration through use, potentially more secure long-term preservation of digital rather than fragile paper), better discovery (through full-text and faceted searching), enhanced usability (mashups, computational analysis, construction of digital corpora), better collection management (through shared storage and metadata management), and more.

Unfortunately, there is no singular process of "digitization" that automatically fulfills all of these promises. Each digitization is unique ([Owens](#)). Every digitization project is the result of a unique intersection of available technologies, budgets, staffing, selected technological standards, and, ideally, the specified needs of a user-community ([Kenney](#)). Indeed, the term "digitization" can be used to describe anything from

photo-imaging that creates eye-readable images of pages, to Optical Character Recognition (OCR) that creates machine-readable text, to the human re-keying of text, to the complete re-creation of a book as an e-book or as an app, to a combination of some or all of these techniques. And digitization – the creation of a digital object from an analog object – is only the first step in delivering on many of these promises. Many other activities (e.g., creating metadata, preserving digital content in a trusted digital repository, providing digital user-services) must also be planned and successfully implemented in order to deliver on those promises.

Although, in an ideal world, the digitization choices for making a book accessible will complement choices for making it preservable, in reality the choices, costs, and technologies for one may actually conflict with those of the other ([Snawder](#)). In the real world, all of those promises may come in conflict with each other, with the intended uses of the resulting digitization, and with the resource constraints of a particular digitization project ([Ames](#)).

In this context, the word "digitization" hides more than it reveals. To know that a book has been digitized does not begin to tell librarians or users what they need to know.

While digitization project managers understand this, it is important that library administrators, collection managers, service providers, preservation officers, business managers, digitization technologists, and others who are responsible for library collections and services also have a clear understanding of what digitization actually delivers, not just what it promises. It is also important for end-users to understand what they are being promised when a library says it plans to deliver its paper collections digitally. This is true not just for libraries that are digitizing their own collections, but also for libraries that intend to rely on digitization done by others. In the digital age, this includes virtually every library since most, if not all, purchase or license access to commercial digital products and rely on publicly available digitizations to complement or supplement their own collections.¹

There are many standards and guidelines for digitization. (See, for example, the *Resource List* created by the [FADGI Still Image Working Group](#)). But these standards are primarily intended for production and are not easily understood or usable either by the end-users who actually *use* the digital copies or by libraries who intend to rely on them. Specifying the technical details of how the object was produced (e.g., bit depth, color space) does not speak to the accuracy of the digitization. Nor does it explain whether the chosen technical specifications were appropriate for any particular use. Indeed, existing guidelines can conflict with each other. They help digitizers make decisions (e.g., file format, pixels per inch), but, ultimately, each project must make choices among complex alternatives. Existing digitization standards do not give us a vocabulary for describing the goals of a digitization project, why the decisions were appropriate, whether or not the project achieved its desired goals, or if the digital objects created meet the needs of end-users.

The existing production standards are essential, but we need to supplement them with a clear, consistent, objective "consumer" vocabulary that non-specialists, including librarians and end-users, can employ to describe and compare the actual utility that digitization brings to users. The DSSOA is one small step toward that end.

Context: Print as an interface

We believe that one base-line, minimum method for the digitization of books is the capturing of the original layout and presentation of the analog work – a digital surrogate for the original analog object.

Books and other printed information packages (e.g., journals, newspapers) don't just store and transport information; they also encode and present information. They *are* the user-interface: the layout of text on the page imparts meaning. Spacing and line breaks define columns and rows in a statistical table and reveal the structure of a poem. The size and position of text provide semantic context that indicates which text is a chapter heading and which is a page number; this context shows that one page contains an original text and the facing page contains its translation. The comparative size and position of text denote footnote numbers and exponential values of numbers. Printed publications use colors, fonts, locations on the page, boxes, lines and other graphic elements to give context to the text (e.g., callouts, sidebars, illustration labels, annotations, footnotes, epigraphs, and headings). Indentation echoes and clarifies hierarchical structure in legal codes and statutes. Variations in font (bold, italics, etc.) can denote emphasis, change in language, and change in narrator voice. Non-text elements contain content: images, photographs, graphics, charts, maps, and so forth. The relative positions of text and non-text elements (position on the page, text and non-text adjacency) is itself information that provides context and specifies inter-relationships between the two. Finally, there is a long history of book design (and design of journals, magazines, newspapers, etc.) which is itself both an art and a science that creates context for the information content ([Bartram](#), [Wikipedia](#)). Text alone without these features can lose meaning or introduce ambiguity and confusion.

As noted above, capturing the original layout and presentation of the analog work is not the only way analog information can be digitized. Nor should capturing the original presentation be the only motivation to digitize. In the long run, this kind of digitization may even be deprecated in favor of the re-creation of content to better match the capabilities and limitations of e-book reading devices and future, as-yet-unimagined technologies. Capturing the original layout is, in some cases, the exact wrong way to deliver content to some e-book reading devices. But creating a digital surrogate is currently a popular method of digitization and it has indisputable, inherent value, which is why we believe it warrants its own user-centered standard.

Introducing the Digital-Surrogate Seal of Approval (DSSOA)

The DSSOA is a tool for asserting that a digital surrogate accurately and completely replicates the content and presentation of a static, analog original such as a book. The DSSOA consists of two Criteria and Rules for applying them.

The Criteria provide a user-oriented, consistent, and easily understood vocabulary for describing the quality of a digital surrogate without the need for specifying the processes or technologies for achieving that quality.

Organizations responsible for the creation, management, preservation, or delivery of digital surrogates may use the Rules to perform the verification of digital objects and then assert that a complete, bibliographically-identified digital surrogate meets the criteria by applying the Seal to it or to a collection of such items.

The DSSOA Criteria may also be used on their own by anyone to describe or compare or evaluate individual titles or entire collections.

Who should use DSSOA

The DSSOA should be used by those curators that seek to replace analog originals with digital surrogates for purposes of preservation or collection management. It should be used when the library's designated user-community expects accurate and complete digital surrogates and when the library expects that users will consult the surrogate instead of the original.

The Digital-Surrogate Seal of Approval

The DSSOA provides a simple tool for doing one simple thing: providing an assurance that the digital object created by a digitization process is as good as the original and that it can be used as easily, as fully, and as accurately as the original. A user given sufficient access to such a digital surrogate would not need to consult the original. The focus of the DSSOA is on the digitization of static, analog originals, such as books, using the definition in ALA's *Minimum Digitization Capture Recommendations* ([Bogus](#)).

DSSOA Criteria

The DSSOA indicates that the original has been reformatted successfully (100% accuracy) and completely. The DSSOA may be applied to a digitized version of an analog original when it accurately replicates the original. To do this, two criteria must be met and verified:

1. **Completeness.** All pages of the original are fully and completely reproduced. No page or part of a page is obscured or blurred or masked or skipped or missing. There is a complete reproduction of each page in its entirety.
2. **Accuracy.** The original layout and appearance are preserved. All text is legible and at least as easily read as the original. There is no visual degradation as compared to the original. All text is clear and without any blur or distortion at the same size as the original. All images are clear and of a quality (e.g., color, resolution) equal to the original.

When both criteria are met according to the rules below, the DSSOA may be applied. If either criterion is not met, the DSSOA may not be applied. Typically, DSSOA will be applied by those that are responsible for the creation or management or preservation or delivery of digital objects.

Rules for applying the DSSOA

The DSSOA may be applied to a bibliographically-identified item when:

1. The responsible organization has verified that both DSSOA criteria are met for that item.
2. The responsible organization provides a Statement of Verification.

The Statement of Verification must:

1. Specify and describe the methodology used to verify compliance.
2. Confirm 100% compliance.

The DSSOA does not specify any method or procedure for verification. Current processes typically use human review of every page ([Rieger](#)), but DSSOA does not prohibit automated verification technologies. The verification statement must specify the methodology used and include either a description of the methodology or a reference to such a description that demonstrates that it successfully verifies human-readable accuracy and completeness. A "digital surrogate" in this context is a complete, accurate, digital

replica of a bibliographically-identified, analog original item.

The Digital-Surrogate Seal of Approval may be applied to bibliographically-identified items or to a collection of bibliographically-identified items when the DSSOA has been applied to all items in the collection.

Use of the DSSOA

The DSSOA criteria may be used by organizations that create, manage, preserve, or deliver digital content. They may apply it to items that they curate as a way of communicating to their user-communities the completeness and accuracy of digital surrogates.

In addition, the DSSOA Criteria may be used by anyone to describe and compare items or collections. For example:

- A digitization project may use the DSSOA criteria to describe intent of the project. Thus, during the planning stages of a digitization project, the planners may assert that they intend to make digitizations eligible for DSSOA.
 - Reviewers of individual items or collections of items may use the DSSOA Criteria to suggest or confirm or challenge a DSSOA assertion.
 - Any user of a digitized item may use the DSSOA Criteria to describe its accuracy or completeness or usability.
-

Notes on applying DSSOA

Although the above description of DSSOA is intended to be unambiguous, it is worth addressing some questions and situations that will likely occur to those wishing to use it.

Digitization. DSSOA can be applied to any process that accurately and completely reproduces in digital form the content and presentation of the original analog item regardless of the technology used to achieve that end. Current technologies used for this purpose produce essentially a photographic image of the original, but future technologies could, presumably, provide the same functionality with different, more advanced techniques.

Statistical Sampling. Checking only a statistical sample of pages to verify a work, or a statistical sample of works to verify a collection does not, by definition, verify every page and so is unacceptable as a method of verifying DSSOA compliance.

Assertion of DSSOA Compliance. DSSOA does not specify any particular format or presentation of the assertion of DSSOA Compliance, but we anticipate that many libraries will create and maintain the assertion in metadata for purposes of description, discovery, preservation, and collection management.

Since most bibliographically-identified items will consist of more than one page or image or digital file, an indication of page-level, image-level, or file-level DSSOA-verification of completeness and accuracy may be asserted in metadata. When all pages or images or files of a bibliographically-identified item have been verified as DSSOA-compliant, the *Digital-Surrogate Seal of Approval* may be applied to the item. The DSSOA may also be applied to a collection of bibliographically-identified items that are all DSSOA-compliant.

Format of Statement of Verification. DSSOA does not specify any particular format or presentation of the Statement of Verification. We anticipate that those asserting DSSOA compliance will do so in metadata that accompany the digitizations. Required information may be included by reference. For example, a single statement that describes the verification methodology used on many items could be indicated by reference in the metadata for each item and the specific date and verifier (e.g., person or process) could be indicated at the page or item level.

Metadata Elements. For libraries that wish to use metadata to identify verification status, we recommend the use of three appropriately structured metadata elements:

- o DSSOA_complete [yes | no ; date = ; verifier= ; notes=]
 - o DSSOA_accurate [yes | no ; date = ; verifier= ; notes=]
 - o DigitalSurrogateSealOfApproval [yes | no ; verification_method= ; qualification= [yes | no]; notes=]
-

Requirements and Exceptions

Missing Pages. The completeness criterion should be interpreted to require the reproduction of any part of the original that conveys context or meaning. This might, for example, include blank pages where such pages provide context (e.g., page numbering, left/right orientation) for the non-blank pages. It might include covers and end-papers. In cases where pages are skipped for justifiable reasons (they provide neither content nor context), they should be specifically documented and justified in the Verification Statement (e.g., "blank end pages contain no content or context and were omitted"). If pages with content or context are intentionally skipped for any reason (e.g., a project lacks the technology for digitizing oversized fold-out pages or maps), the bibliographically-identified item cannot be considered DSSOA-compliant.

Legibility of text. Digitized text that is blurred, or less sharp and clear than the original at the same size as the original, disqualifies an item from getting the Seal. Anything that makes the digitization less readable or less legible than the original, in whole or in part, would disqualify the entire item.

Artifact. With current technologies, digital surrogates do not normally provide a reproduction of the original artifact. DSSOA therefore does not include any criteria for evaluating a replication of the artifact ([Council on Library and Information Resources](#), [Werner](#)).

Qualified DSSOA-Compliance. Incomplete or damaged original. DSSOA may be applied to a complete digitization of an incomplete or damaged original where no complete or undamaged original exists. In such cases, the DSSOA criteria must be applied to what is available for digitization and the Verification Statement must record that the surrogate is an accurate and complete copy of an incomplete or damaged original. DSSOA-compliant items may then use the label "Qualified Digital-Surrogate Seal of Approval: some content [may be | is] damaged or missing."

Qualified DSSOA-Compliance. Digitizations of derivatives. When a digitization is created from a derivative (e.g., digitization of a microfilm copy of a newspaper), this must be clearly recorded in the Verification Statement and any known problems with the derivative (i.e., if the derivative itself would not meet the DSSOA criteria -- for example if the completeness or accuracy of the derivative is not verified or if accuracy of the derivative has known problems such as missing pages or inaccurate reproduction of illustrations)

must be indicated. DSSOA criteria must be applied to what is available for digitization. DSSOA-compliant items may then use the label "Qualified Digital-Surrogate Seal of Approval: some content [may be | is] damaged or missing."

DSSOA collections. When a collection of bibliographically-identified items all have the same, unqualified Digital-Surrogate Seal of Approval, the collection may use the designation "Digital-Surrogate Seal of Approval" to describe the collection.

Qualified DSSOA collections. If some or all items in a collection have only Qualified DSSOA-compliance (see above), the collection may only use the designation "Qualified Digital-Surrogate Seal of Approval: some copies are damaged or incomplete."

Newly discovered problems. If, after DSSOA is applied, an inaccuracy, illegibility, or incompleteness is discovered, either the DSSOA assertion must be removed, or the problem corrected and re-verified. In the case of an item in an DSSOA collection, the work must be removed from the collection, or corrected, or the collection must be down-graded to be no longer DSSOA compliant.

Value of the DSSOA in a Larger Context

The DSSOA will not be appropriate or necessary for all digitization projects or collections. It should, however, be of value to those libraries that wish to use digitization as a means of collection management ([Schonfeld](#), [Lavoie](#), [Malpas](#)). This is particularly true when a library plans to use the digital object as a substitute for rather than as a complement to the original analog copy. It should also be of value to libraries that intend to rely on the digital collections of commercial or non-commercial organizations for delivery of content and services to their users. In such environments, DSSOA provides an essential metric for evaluating options and for communicating to users.

The interplay between digitization, collection management, preservation, and service delivery is complex and no single standard can provide adequate guidance on its own for library decision making. We believe that libraries in the digital age will need a set of tools that will facilitate wise long-term decision making and that the DSSOA is only one such tool. We anticipate other tools that will provide a clear way of communicating with users such things as:

- Quality, kind, and extent of discovery metadata and tools.
- Long-term preservation guarantees including deposit in trusted repositories.
- Usability and utility including APIs.
- Selection criteria including accuracy, currency, coverage, etc.

Digitization is just one aspect of a rapidly changing information environment. With tools like DSSOA, libraries should be better equipped to communicate to their users the value of the library in a clear, unambiguous way.

Summary

The *Digital-Surrogate Seal of Approval* is a tool for asserting that a digital surrogate accurately and completely replicates the content and presentation of a static, analog original such as a book. It should be used by those curators that seek to substitute a digital surrogate for an analog original for purposes of

preservation or collection management and for access purposes where accuracy and completeness are required by the user-community.

The DSSOA Criteria provide a user-oriented, consistent, and easily understood vocabulary for describing the quality of a digital object without the need for specifying the processes or technologies for achieving that quality. These criteria may be used by anyone to describe or compare digital surrogates or collections.

Notes

¹ There are many such examples of publicly available digital collections but a few of the most popular ones include the [Internet Archive](#), [Google Books Project](#), [Hathitrust](#), Library of Congress' [American Memory Project](#), [University of North Texas Digital Library](#), and the many projects listed on the [Federal Depository Library Program Digital Registry](#).

Endnotes

[1] Ames, Eric. "[So We Can Throw These Out Now, Right?: What We Learned From Microfilming Newspapers and How It Shapes Our Digitization Strategy](#)." *The Baylor University Libraries Digital Collections Blog*. August 23, 2012.

[2] Bartram, Alan. 2001. *Five hundred years of book design*. New Haven, CT: Yale University Press.

[3] Bogus, Ian *et al.* 2012. [Minimum Digitization Capture Recommendations \(DRAFT\)](#). Association for Library Collections and Technical Services Preservation and Reformatting Section.

[4] Council on Library and Information Resources (CLIR). (2001). [The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections](#) (publication No. 103). Washington, D.C.

[5] FADGI. Still Image Working Group. [A Resource List for Standards Related to Digital Imaging of Print, Graphic, and Pictorial Materials](#). Federal Agencies Digitization Guidelines Initiative. January 28, 2010.

[6] Kenney, Anne R. "[Digital Benchmarking for Conversion and Access](#)." In Kenney, Anne R., and Oya Y. Rieger. *Moving Theory into Practice: Digital Imaging for Libraries and Archives*. Research Libraries Group, 2000. p.24-60.

[7] Lavoie, Brian F., Constance Malpas, and J. D. Shipengrover. 2012. [Print Management at "Mega-scale": a Regional Perspective on Print Book Collections in North America](#). Dublin, OH: OCLC Research.

[8] Malpas, Constance. 2011. [Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment](#). Dublin, Ohio: OCLC Research.

[9] Owens, Trevor. "[All Digital Objects Are Born Digital Objects](#)." *The Signal: Digital Preservation*. (May 15, 2012).

[10] Rieger, Oya. 2008. "[Preservation in the Age of Large-Scale Digitization: A White Paper](#)."

[11] Schonfeld, Roger C., and Ross Housewright. (2009) [What to Withdraw: Print Collections Management in the Wake of Digitization](#). Ithaca S+R.

[12] Snawder, Kristin. "[Digitization is Different than Digital Preservation: Help Prevent Digital Orphans!](#)"

The Signal: Digital Preservation. (July 15th, 2011).

[13] Werner, S. (2012). [Where material book culture meets digital humanities](#). Presented at the Geographies of Desire, University of Maryland.

[14] Wikipedia. "[Canons of page construction](#)."

About the Authors



James A. Jacobs is Librarian Emeritus, University of California San Diego. He has more than 20 years experience working with digital information, digital services, and digital library collections. He is a technical consultant and advisor to the Center for Research Libraries in the auditing and certification of digital repositories using the Trusted Repository Audit Checklist (TRAC) and related CRL criteria. He served as Data Services Librarian at the University of California San Diego from 1985 to 2006 and co-taught the ICPSR summer workshop, "Providing Social Science Data Services: Strategies for Design and Operation" from 1990 to 2012. He is a co-founder of [Free Government Information](#).



James R. Jacobs is the Federal Government Information Librarian at Stanford University's Cecil B. Green Library and program lead for the [LOCKSS-USDOCS](#) program. He is a member of the Government Documents Roundtable (GODORT) of the American Library Association and has served on Depository Library Council to the Public Printer, including as DLC Chair from 2011 - 2012. He is co-founder of Free Government Information and [Radical Reference](#) and serves on the board of [Question Copyright](#), a 501(c)(3) organization that promotes better public understanding of the history and effects of copyright, and encourages the development of alternatives to information monopolies.

Copyright © 2013 James A. Jacobs and James R. Jacobs

PRINTER-FRIENDLY FORMAT

[Return to Article](#)
