

November 2013

## Future Tense -- The Disapproval Plan: Rules-Based Weeding & Storage Decisions

Rick Lugg

*R2 Consulting LLC*, [rick@r2consulting.org](mailto:rick@r2consulting.org)

Ruth Fischer

*R2 Consulting LLC*

Follow this and additional works at: <http://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

---

### Recommended Citation

Lugg, Rick and Fischer, Ruth (2008) "Future Tense -- The Disapproval Plan: Rules-Based Weeding & Storage Decisions," *Against the Grain*: Vol. 20: Iss. 6, Article 32.

DOI: <https://doi.org/10.7771/2380-176X.2627>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

# Future Tense — The Disapproval Plan: Rules-Based Weeding & Storage Decisions

by **Rick Lugg** and **Ruth Fischer** (R2 Consulting LLC, 63 Woodwell's Garrison, Contoocook, NH 03229; Phone: 603-746-5991; Fax: 603-746-6052) <rick@r2consulting.org> [www.r2consulting.org](http://www.r2consulting.org)

Credit where credit is due. Much as we'd like to claim to have originated the phrase "disapproval plan," it happened like this. During a workflow analysis project at **Davidson College**, R2 had written this recommendation: "**Adopt a rules-based approach to weeding monographs.**" In explaining it, we suggested that the Library collaborate with teaching faculty (who have the final say on selection and de-selection at **Davidson**) to define categories of books that could be withdrawn without title-by-title review, enabling a batch approach to some weeding decisions. As we elaborated on the idea for **Jill Gremmels**, Director of the library at **Davidson**, we characterized it as a sort of "reverse approval plan." To which she replied with a laugh, "Oh, you mean a *disapproval plan!*" We know a good moniker when we hear one.

Beyond its excellent (if slightly Puritanical-sounding) name, the disapproval plan is one of those ideas whose power seems immediately obvious. In our September "**Future Tense**" column in *Against the Grain*, we outlined the reasons that now is the time to usher in a **Golden Age of Weeding**. In this article, we present a new technique that R2 is developing to hasten this golden age along. While the full systems and service package we envision will actually be called **Sustainable Collections Service (SCS)**, parts of it will remain fondly known as the "disapproval plan." R2 believes that many elements of approval plan profiling can be adapted to much-needed de-selection and de-accessioning processes. As with selection, these techniques can help make de-selection more efficient while assuring that collection integrity is maintained.

## The Approval Plan

Since the early 1970s, many libraries have employed a rules-based approach to selection and collection building: the approval plan. No library uses approval plans to select all of its material, but in most a set of rules (the profile) operates against a defined universe of newly published titles, generating a combination of books, notifications, and exclusions. Although there is sometimes selector review of individual titles directly shipped to the library, this activity typically diminishes once the profile has matured — i.e., once confidence in the rules has been established. Selector efforts are directed toward refining the profile, rather than toward consideration of individual titles.

Although methods vary somewhat from vendor to vendor, profile "rules" typically consist of three or four components:

**Defining the Universe:** In determining what will be covered in their approval program, some vendors work from specific publisher lists, which are revised slightly from year to year. Others shape coverage from national bibliographies, based on the judgment of their staff. This universe serves as a group of "candidate" titles for selection and acquisition.

**Selection Metadata – Non-Subject Parameters:** Over time, vendors and librarians have evolved sophisticated vocabularies to categorize published content. These include publication types (reprints, textbooks, Festschriften); series type (numbered, unnumbered, annuals), content and readership levels, format descriptors (paper/cloth; print/electronic; reference types), alternate editions, and many others. This metadata is applied by vendor staff, based on direct examination of the book or an electronic surrogate.

**Selection Metadata – Geographic, Historical, and Interdisciplinary Descriptors:** Well-crafted book descriptions (and rules that operate against them) extend beyond subject classification to identify geographic or historical focus, and to accommodate subjects that cross disciplinary boundaries (e.g., Women's Studies), or that highlight specific facets of a

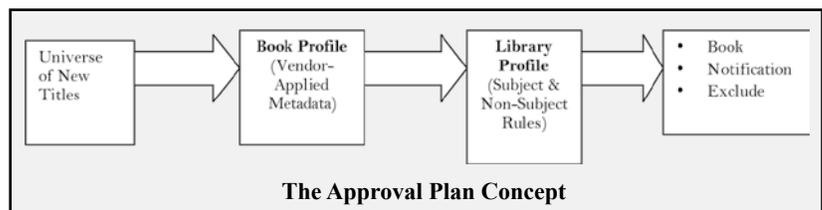
topic (e.g., public policy aspects of Medicine).

**Selection Metadata – Subject Classification:** Most vendors support subject description based on the major classification schemes: LC, DDC, NLM, and/or their own subject thesaurus. Content is described in accordance with library practice, and most titles are classified in more than one of these schemes, to enable the vendor to support profile rules that match the individual library's approach.

**Library Profile:** Typically, the library uses the same vocabulary of selection metadata to describe its collecting interests and priorities — the rules to be applied to the universe of newly published content.

**Application of Rules:** Each of the selection metadata components is used in two ways: 1) to describe the content being considered (the book profile); and 2) to describe the rules that selectors want applied to candidate titles (the library profile). Each week, new titles entering the vendor system are described and then compared to library profiles (sometimes by vendor staff, sometimes by automated decision support systems, sometimes by a combination of the two). This process results in a decision to take one of three actions: send book, send notification, or exclude.

The rules that govern approval plans can be quite detailed, and can be developed in many ways. The best vendor representatives have mastered the variables of their company's approach, and can suggest when separate plans are advisable, or how to use sub-profiles to segregate reference material, or when to vary non-subject parameters in specific subject areas. After more than 30 years of refinement, approval plans are widely accepted as an efficient method for selecting and acquiring mainstream content. When supported by batch copy cataloging, electronic invoicing, and shelf-ready services, a well-maintained approval plan can allow selectors and acquisitions staff to turn their attention to other priorities.



## Current Weeding Practice

Current weeding practice in most libraries (when it is done at all) is to have selectors review candidate titles one by one; i.e., the way that most *selection* in libraries occurred before the advent of approval plans. This is a daunting task, even in the unlikely event that sufficient selector time is available, or that the task is accorded a high enough priority. This seems an ideal situation for the application of a rules-based approach. The material to be considered is by definition low-use or no-use material. But libraries, especially research libraries, have been built on a premise that current use (or lack of use) alone is an uncertain determinant of scholarly value. Research libraries support a policy of "just in case" availability, and a conservative approach to de-selection. After all, how can we know what future scholars may need? How can we risk discarding any content? There is a powerful underlying fear of making an irreversible decision — of throwing out something that may later be wanted, although it appears to be of minimal value now. What if no library holds the exact work that's wanted?

There is some validity to these fears, given our roles as caretakers of the scholarly record. At the same time, a walk through the stacks in

*continued on page 75*

any academic library makes plain that there are plenty of opportunities to weed without beginning to threaten the library's mission. And, of course, not all libraries are research libraries.

In our networked, digital world both risks and opportunities are changing. Under an access (rather than ownership) model, the key question becomes "how likely and at what cost can I access this again?" rather than "is it in my library's collection?" For millions of titles, full-text digital surrogates exist. For millions of others, ILL or resource sharing arrangements can be relied upon for retrieval. (This assumes that we as a community take a coordinated approach to weeding, sharing last-copy responsibility, and assuring that no content disappears completely.) The Web has made it possible to discover and obtain used copies of millions of other titles.

### The Disapproval Plan

The disapproval plan, known more formally as **R2's Sustainable Collections Service (SCS)**, takes as its premise that similar rules-based techniques can be applied to weeding and storage decisions. In brief, a library-defined de-selection profile operates against a candidate file of low-use titles, generating a provisional decision for each title: withdraw, store, or retain. The SCS methodology [patent pending], consists of the following components:

**Defining the Universe:** The first step is to determine the parameters that make a title eligible for consideration for withdrawal. Most automated library systems can generate a list of titles that meet specified inactivity thresholds; e.g., circulating monographs with an imprint date of 1990 or earlier that have not circulated within the past ten years. These lists of "candidates" for weeding or storage provide the starting point for de-selection decisions.

**De-Selection Metadata – Non-Subject Parameters:** The vocabularies that categorize content into publication type (reprints, textbooks, Festschriften); series type (numbered, unnumbered, annuals), content and readership levels, format (paper/cloth; print/electronic; reference types), when available, can also be used effectively for de-selection. Some format and edition-related metadata (including the existence of alternate editions) can be gleaned from the bibliographic record; other information can be obtained from links to extended metadata. While this metadata may not be as complete for older titles, the book itself is of course available for examination.

**De-Selection Metadata – Geographic, Historical, and Interdisciplinary Descriptors:** Subject headings, classification, and 5xx content notes, and other fields can be used to identify a work's geographic or historical focus, and to accommodate subjects that cross disciplinary boundaries (e.g., Women's Studies) or classifications, or that highlight specific facets of a topic (e.g., public policy aspects of Medicine).

**De-Selection Metadata – Subject Classification:** Since the candidate file consists of (or links to) MARC records from the Library's ILS, subject classification will be present for most titles. This serves as the basis for the book profile, which must then be supplemented with de-selection metadata.

**Library Profile:** As with an approval profile, selectors would decide how to manage de-selection in their disciplines, based on an overall strategy set by the library administration. In de-selection, the library profile would specify the library's level of tolerance for risk and potential cost if re-access is needed. Risk and cost tolerances might vary by subject, format, imprint date, or other factors. (These are further described below.)

**Application of Rules:** As with approval plans, the SCS metadata elements would be used in two ways; 1) to describe the content under consideration (the book profile); and 2) to specify rules to apply to low-use candidate titles (the library profile). On a schedule determined by the library, candidate titles would be augmented with de-selection metadata, and compared to the library's "disapproval" profile via the SCS decision-support system. The process would result in one of three provisional actions: retain, store, or withdraw.



Rules governing de-selection may be as simple or detailed as wanted, and would vary by subject or perhaps other factors such as space needs in a given range of stacks. When supported by batch maintenance of item and bib records, this approach provides a much more scalable and efficient tool for managing de-selection. In R2's view, rules-based weeding could become a reality with the introduction of three new concepts for de-selection metadata. We call these the **Surrogate Collection Index**, the **Withdrawal Risk Factor**, and the **Access Cost Factor**.

**Surrogate Collection Index:** To identify available surrogates for the content under consideration for withdrawal, R2's SCS system would interact with a variety of target databases:

- **OCLC WorldCat:** to identify other libraries holding the same item
- **Google Book Search, the Internet Archive, and the Million Book Project:** to identify full-text electronic versions of the same content in the public domain
- **Commercial eBook providers:** to identify where the title might be available for rental or re-purchase in electronic form
- **Lightning Source** and other print-on-demand providers: to confirm whether a candidate title is available via POD
- **Alibris, Amazon Marketplace, and other Web booksellers:** to gauge the extent of availability of used copies of the same title
- **Amazon Historical Pricing:** to prevent inadvertent discard of valuable titles

Based on SCS interrogation of these sources, R2 would establish two key measures for each title:

**Withdrawal Risk Factor (WRF):** A numerical score that indicates the potential difficulty of re-accessing or re-purchasing withdrawn content in the (unlikely) event that it is subsequently wanted. The lower the WRF, the more confident the library can be about discarding the title or copy. The higher the WRF, the more likely it would be retained or stored.

Withdrawal Risk Factor	Examples
0 – No Risk	<ul style="list-style-type: none"> <li>• An identical print version is held by 10 WorldCat libraries or 3 resource sharing partners</li> <li>• Title remains widely available for re-purchase</li> </ul>
1 – Minimal Risk	<ul style="list-style-type: none"> <li>• An identical or acceptably similar version has been digitized in full text by Google Book, Open Content Alliance, or similar sources; content is in public domain</li> <li>• Title is out-of-print but available via commercial eBook provider(s).</li> </ul>
2 – Moderate Risk	<ul style="list-style-type: none"> <li>• Variant edition is held by other libraries, available via ILL</li> <li>• Identical or acceptably similar version digitized by Google or OCA, but content is under copyright</li> </ul>
3 – High Risk	<ul style="list-style-type: none"> <li>• Identical or substantially similar version is not held in WorldCat</li> <li>• Available only in another language</li> <li>• Available only in part</li> <li>• No acceptable surrogate exists</li> </ul>

**Access Cost Factor (ACF):** A score that rates the potential cost (in both staff time and cash outlay) of re-accessing or re-acquiring a title. The higher the ACF, the more conservatively a library might act in discard decisions — while bearing in mind that these titles have not been called for in 20 years!

*continued on page 76*

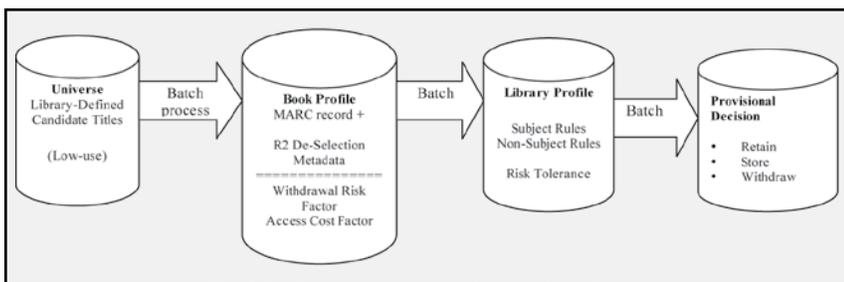
Access Cost Factor	Examples
0 – No Cost	<ul style="list-style-type: none"> <li>No such animal</li> </ul>
1 – Low Cost	<ul style="list-style-type: none"> <li>Available via ILL or direct borrowing</li> <li>Identical or substantially similar version available via Google Book Search or Internet Archive (public domain titles only)</li> </ul>
2 – Moderate Cost	<ul style="list-style-type: none"> <li>Identical or substantially similar version available via commercial eBook providers (ebrary, EBL, NetLibrary, MyiLibrary, publisher)</li> <li>Print-on-demand version available from Lightning Source</li> <li>Used copies currently available on Web market (Alibris, abebooks)</li> </ul>
3 – High Cost	<ul style="list-style-type: none"> <li>Title out-of-print; available only at a premium</li> <li>Title has never been digitized</li> <li>Available in Special Collections only; restricted resource sharing</li> </ul>

An implementable system would, of course, be much more refined than these examples. The **WRF** and **ACF** would be assigned independently of one another, but clearly their combined effect would need to be considered. A low-risk/high-cost item differs from a low-risk/low-cost item. Weighting of factors and additional rules would be needed. Definitions for “acceptably similar” and other terms will also need to

be developed. Flexibility would be critical across varying subjects. But even these simple outlines show the potential power of using proven techniques to solve a growing problem. And just in time.

We envision that **WRF** and **ACF** would become elements in both content description and in the rules that operate against candidate titles. Running library-defined rules against the candidate file will result in a numeric “score” for each item. A high score would suggest high risk and/or high cost, and retention as a provisional action. A middle score would suggest storage, and a low score potential withdrawal.

The arguments that ultimately resulted in the widespread acceptance of approval plans to support selection are equally valid when applied later in the life-cycle of book content. Profiles and rules assure consistent treatment of all subject areas — and of the collection as a whole. These tools can also help assure collection integrity while pursuing even very aggressive weeding and storage targets. Finally, an approach to de-selection that is based on rules, batches, and automated support enables many titles to be handled efficiently. The end result: a high-volume, high-integrity solution to moving low-use titles out of the main library — opening additional space for users. 🌱



The SCS (Sustainable Collection Service) Concept

*Please note: this article specifically addresses the weeding of monographs from circulating collections. There exist many other weeding opportunities for print serials, Government Documents, print reference collections, micro-formats and audio-visual material. These will be addressed in subsequent articles in this series. Our next article will take up the issue of “Sustainable Collections: Maintaining the Library’s Carrying Capacity.”*

## Issues in Vendor/Library Relations — Buying eBooks: Does Workflow Work? Part II

by **Carolyn Morris** (Director of New Business Development, Coutts Information Services) <cmorris@couttsinfo.com>

Column Editor: **Bob Nardini** (Group Director, Client Integration and Head Bibliographer, Coutts Information Services) <bnardini@couttsinfo.com>



Choosing an eBook vendor is hard. That’s because to a certain extent it requires predicting the future — the future of the business, of the vendor, even the future of the format itself. Rather than guess wrong, it can seem safer to hold off; but if decision makers are thoughtful, they can make choices now that will position their libraries well not only for today, but also for the future.

Since eBooks do not live on the library’s shelves it is important, first and foremost, to build eBook collections with a company that will be around for a long time, and able in the first place to stay committed to eBooks. Though it is impossible to know for sure which companies will thrive in the future, it does make sense to examine track records and ask

questions about business models. Are eBooks a core part of the business or just a sideline? Does the company have a history of pulling the plug if profits aren’t quickly forthcoming? Is the company offering a deal that seems too good to be true? Is the business model viable? In short, smart libraries will choose to work with companies that are making smart business decisions now.

Content is another thing to evaluate. Currency, quality, breadth, and relevance to academic libraries all are important, and patrons benefit from having as much good content aggregated on a single platform as possible. Though some users can learn a new interface without too much effort, many, and probably most, have no understanding of the library’s

digital infrastructure. Multiple platforms require repeating the same search several times to uncover all relevant resources, a fact unknown and likely of little interest to many users. The more platforms, the more likely a researcher will miss useful eBooks. The **OPAC**, of course, sometimes will allow patrons to search across all of the library’s holdings, but discovery is limited to the elements of the **MARC** record. The value of a stronger search that would lead a user to more of a library’s digital content is lost.

There is no way to know for sure which eBook providers will be most successful in aggregating desirable content, but there are indicators of likely future success. The quan-

*continued on page 77*