

E-Analytics: Issues and recommendations for processing electronic collections and series – Final Report

Date: November 30, 2010

Authors: Kirsten Clark, Chew Chiat Naun, Cecilia Genereux, Lisa Johnston, and Stacie Traill

The authors acknowledge the work of the *E-analytics Subgroup of the Cataloging Coordination Group (CCG)*, which included the authors, and Nicole Theis-Mahon and Christopher Herzberg from Bio-Med.

Table of Contents

1. Summary of recommendations
2. Introduction
3. Benefits to users
4. Issues
 - 4.1. What to Catalog
 - 4.1.1. Discoverability
 - 4.1.2. Ongoing access/stability
 - 4.1.3. Stewardship
 - 4.1.4. Free vs. paid resources
 - 4.1.5. Availability of records
 - 4.2. Sources of records
 - 4.2.1. Batch loading of pre-existing record set
 - 4.2.2. Batch loading of locally generated record sets
 - 4.2.3. Manual copy cataloging
 - 4.3. Workflow
 - 4.3.1. Identifying new content
 - 4.3.2. Collection maintenance
 - 4.3.3. Tracking treatment decisions
 - 4.3.4. Projects vs. regular work
5. Recommendations
 - 5.1. Criteria for analytic treatment
 - 5.2. Sources of records
 - 5.2.1. Brief records
 - 5.2.2. Vendor services
 - 5.3. Workflow and communication
 - 5.3.1. Augment the e-resources from
 - 5.3.2. Document requests and their status
 - 5.3.3. Notification and monitoring
 - 5.4. Timely decisions on cataloging requests
6. Conclusion and Next Steps

Appendices

1. Example Case Studies
2. Meetings with Academic Program Departments
3. Background
4. ERL Questions for selectors and vendors
5. Pilot Project

1. Summary of recommendations

- Priority for analytic treatment of a title should be given on the basis of distinctiveness, non-duplication of access, stability, stewardship, availability of records, and cross-format access.
- A change from print to electronic distribution for a title should trigger a review of cataloging treatment.
- Consider using brief records for some classes of material.
- Work with YBP to receive notification and cataloging for new additions.
- Technical Services and Academic Programs should work to clarify workflows and responsibilities.
- Augment the e-resource form with additional information to inform the analytic treatment of a resource.
- Document all cataloging requests for online series and collections.
- Establish a process for monitoring new titles, including identifying and registering for alerts.
- Review and make timely decisions on all cataloging requests. Projects should be reviewed by Technical Services and Collections Council.
- Don't wait for a complete solution to all e-analytic problems. Begin by looking for quick wins.

2. Introduction

***Definition:** The term “e-analytics” is used broadly here to describe title-level cataloging of series and collections that are wholly or partly available online. Examples include monographic book series and online reports. They may be paid or free. The use of the word “analytics” in this context does not necessarily imply the presence in the catalog of both a series-level and a title-level record.*

The problems raised by e-analytics are the direct consequence of the differences between print and online distribution. In the print world, the selector purchased a serial or monographic series, and the decision whether to analyze the incoming item at title-level was usually made within Technical Services based on established criteria. Often, the cataloging practice was established on the first volume received and ongoing additions to the series were added as each physical item was received. The physical volume itself served as the channel of communication in the workflow.

In the digital world, the decision to analyze and workflow for ongoing additions are not as clear. There are no physical volumes to arrive and catalog. The current procedure of adding electronic series to the catalog involves selectors and Collection Development and Management staff making a request, via email, to Technical Services staff. The request is handled on case-by-case basis depending on title and information provided with no ongoing maintenance of title or website. Broader issues are also at work here, such as the availability of alternative discovery paths and the longevity (or otherwise) of born-digital items. Web publishing is diverse with different formats and different levels of stable access.

In 2009, the E-analytics Working Group was tasked by the Coordinated Cataloging Group to look at the overall picture of e-analytics, determine what the issues are, outline specific types of case studies, and offer recommendations for solutions to the issues, to the extent that staffing capacity and overall priorities allow.

The group was given four priorities to look at:

- Identify categories of monographic and serial online materials in need of cataloging treatment,
- Propose solutions for providing access to materials not adequately covered by existing procedures,
- Estimate the resource implications of these solutions, and
- Propose a set of policies for deciding treatment

During the Fall of 2009 and Spring of 2010, the E-analytics Working Group outlined the issues of e-analytics, met with Academic Program divisions, gathered examples, and discussed recommendations. Appendix 1 of this report provides a representative list of electronic collections that we investigated. Appendix 2 summarizes the group's discussions with Academic Programs. Appendix 3 provides some background on issues pertaining to the cataloging of analytics.

3. Benefits to users

While this report focuses on the aspects of e-analytics from the internal library standpoint, the current work and future endeavors in this area ultimately focus on how the end user will be benefited by changes to the current policies and procedures. While by no means extensive the following points have been raised regarding users' use of e-analytics.

- Cataloging e-analytics provides better user access for those titles that are not easily found in a typical Internet search engine result list.
- No one discovery mechanism fits all types of research so may be necessary to provide multiple access points to ensure adequate availability.
- Good stewardship of our collection and resources requires that we provide the best access possible. It does not make sense to spend staff resources and in some cases purchasing resources when there is limited or non-existent access.
- Stewardship also extends to the future and ensuring access to resources for future generations of researchers.

4. Issues

4.1 What to catalog

The most important question in dealing with e-analytics is, of course, what to catalog. The following factors are usually taken into account when making decisions about analytic treatment.

4.1.1 Discoverability

The local library catalog has traditionally been the main discovery environment for print monographs and serials. By contrast, users have long relied on a variety of other tools to discover online resources. In recent years the trend has been not to analyze materials that are readily discoverable elsewhere.

Some of the considerations applied in deciding whether to analyze print serials are also relevant to online resources. Following its 2003 review (see Appendix 3) Technical Services stopped analyzing serials that were indexed with analytic-level metadata in abstracting and indexing databases such as Academic Search Premier, Applied Science and Technology Abstracts, AGRICOLA, and ISI Web of Knowledge. The Health Sciences Library ceased analyzing many serial titles and now relies on PubMed to make electronic content discoverable at the analytic level.

It is also a well-established principle in Technical Services that analytics are warranted only when the analyzed volumes have distinctive content that is not adequately captured by a collective or series-level record. True monographic series, where the monographic author and title are more prominent than the series title, are strong candidates for analytic treatment. On the other hand, quasi-monographic, journal-like series, where the series title is more prominent than the monographic title, generally do not warrant analysis.

Nevertheless, there are cases where users are clearly better served by being able to find analyzed content in the library catalog. Collections like the *Lecture Notes in Computer Science* were analyzed in print and continue to be provided with analytic-level records through batch loads. Maps that were once printed and delivered to the library from state geological surveys are now consistently published online, with a print-on-demand option to reduce costs. Subsequently, users who are looking for a map in the catalog will only find the older, out-of-date version of this research.

Recent developments in library discovery environments widen the range of available options. It is possible, for example, to pipe records into the Libraries' Primo environment without first creating records in Aleph. Taking this approach has some potential to simplify workflows and allow more flexibility in the type of cataloging provided. A potentially significant development is the introduction of network-level library discovery environments such as WorldCat Local, Serials Solutions' Summon, and Ex Libris' Primo Central. In these environments records for widely held online resources can be maintained centrally, increasing the incentive for providers to supply records and greatly reducing the amount of maintenance that individual libraries need

to perform.

4.1.2 Ongoing access/stability

Resources that are known to be stable and to which the Libraries have secured perpetual access are given priority for cataloging over resources for which no such assurances are to be had. There are good reasons for this policy. A title that has been acquired only on leased access, or which comes as part of a package from which it may at some point be dropped, may have only a short life in the catalog and it becomes more difficult to justify the expense of cataloging it. In addition to this initial expense of cataloging, it imposes a considerable burden to monitor links and maintain links in the catalog. In the case of e-journals, a solution has evolved in the form of Ex Libris' SFX OpenURL resolver and its complementary MARCIt service, which eliminate local, manual updating of links for thousands of electronic journals. But no corresponding service exists for most monographs.

4.1.3 Stewardship

In some instances a strong case can be made that the Libraries have a role in ensuring continued access to potentially ephemeral web resources. For example, the Water-Resources Investigations Report, Minnesota (sys. no. 000111494) were received in print for the last half-century but moved to online-only in the early 2000's and there is some doubt over their persistence. These data, valuable in important areas such as climate change, could be left incomplete for future generations. Such titles still require stewardship for future users, regardless of current web access. In these cases there is arguably a concomitant responsibility to make sure that the resources are adequately identified and described within a stable discovery environment.

Issues concerning the archiving of digital content are beyond the scope of this report, but we note here that stewardship can have an access component. For example, it may be important to not only to describe a resource but also to identify its authors accurately, or to make clearly evident the organization or collection that it derives from. In some cases it can also be important to identify a specific version of the resource unambiguously, or to relate it to other versions.

4.1.4 Free vs. paid resources

Historically materials acquired through purchase or subscriptions have received higher priority for cataloging than free online resources. Free resources tend to be both more openly discoverable and less stable, and therefore often score lower on other criteria for analytic treatment. However, it is not clear that *all else being equal* there are any reasons to treat free resources differently from paid ones.

4.1.5 Availability of records

The availability of records is an important consideration when dealing with large batch loads of electronic items acquired as collections, such as thousands of Springer E-books. The availability of record sets is not as important when selecting a single item or monographic serial. While the

availability of MARC records should not be the deciding factor for selection for cataloging, it obviously has a large bearing on the feasibility of providing catalog access. In some cases records are available at additional cost. In these cases the cost versus benefit of acquiring the records need to figure in purchasing decisions, along with responsibility for payment.

Equally important are the workflows that a given vendor will support. It is a significant advantage, for instance, if updates are provided on a timely basis and the records contain match points to facilitate maintenance.

4.2 Sources of records

The availability of MARC records, or credible information from which MARC records could be automatically generated, is an important consideration in e-analytic treatment decisions. Some series or collections may have records freely available for batch loading, but this has not generally been the case for free resources. The options for catalog records, in order of preference, include the following.

4.2.1 Batch loading of pre-existing record sets

Some series have analytic records because they are part of electronic book collections for which vendors or publishers provide sets of bibliographic records. These records are added to the catalog via batch loads. Batch loading is a well-established, well-understood workflow. When such records are available, this is the obvious choice for providing title-level access to monographic resources in collections or series. Alternatively, other institutions may have made formal commitments to catalog a particular collection or series. It is possible that those institutions might be willing to extract and share their MARC records in such cases.

Volumes for series may also be ordered singly through GOBI, or specific series may be added to relevant approval plans. Titles purchased through GOBI automatically receive a cataloging record from YBP, whether they are print books or e-books. These records are loaded into the catalog as part of a regularly scheduled batch process managed by Enterprise Technology staff.

4.2.2 Batch loading of locally generated record sets

While this is something we have done very little of in the Libraries, other institutions have successfully batch loaded locally generated records sets for some resources. Such workflows presuppose the availability of a spreadsheet or other formatted list of titles in a collection/series that provides basic access information such as ISBN, title, and URL. While a program such as MARCEdit has some ability to generate records from this type of data, it is also possible that programming time and expertise would be required for the Libraries to pursue this strategy. We believe the requisite programming expertise exists in the IT Division, but arrangements would have to be made to allow the appropriate individuals to take the time to complete such projects.

It should also be noted that this process would produce very minimal records, and a relatively high error rate is possible (depending on the reliability of the information provided by from

vendors/publishers). But if the goal is simply basic title-level access, such records may be sufficient.

Another alternative is to batch import records from OCLC. This approach might work especially well for sets or collections where copy is almost always available soon after publication. Adopting such a workflow would necessitate training for Technical Services staff, and collaboration with Enterprise Technology to develop loader mapping for the creation of item and holdings records.

4.2.3 Manual copy cataloging

Records are imported to Aleph one at a time from OCLC, then edited as needed in Aleph. While we have refined this process to make it as efficient as possible, it is not a reasonable option for large collections/series (e.g., those with thousands of titles). If we were to take on manual cataloging of more e-resources, additional staffing would be required (along with training). The manual cataloging approach is also the most difficult in terms of ongoing maintenance, since most collections/sets do not offer an easy way to track updates. This is in contrast to batch loads, where in many cases, vendors/publishers provide regularly updated record sets that can be loaded on a schedule.

Another consideration is that not every title is likely to have copy in OCLC. The need to route such titles for original cataloging means the cost of cataloging such titles becomes much higher. Minimal level cataloging for such titles may be an alternative worth considering when manual cataloging is the chosen option for cataloging a collection/series.

4.3 Workflow

Workflows for e-analytics are more complex than those for print resources. With print resources the arrival (or withdrawal) of the physical piece served as the trigger for the relevant actions in Technical Services. With electronic titles other mechanisms must be found, and they will require clear channels of communication and allocations of responsibility between Academic Programs and Technical Services.

Currently information about new resources follows two main paths: if it is a resource that is purchased, it goes through the Electronic Resources Librarian, or if it is a free resource and the web resource form is filled out, it goes through catalogers in Technical Services.

4.3.1 Identifying new content

Timely awareness of the availability of new content is one of the most problematic aspects of providing title- or issue- level cataloging. Many publishers lack any kind of notification service. Where notification is available, it can take several forms, few of them easy to integrate into a cataloging workflow. The methods in use include:

- Email
- RSS feeds
- New content is posted to a web page only with no information pushed out to subscribers

- New content is added to existing content lists with no indication about which titles/issues/volumes are actually new.

The frequency of the electronic content updates varies as well, as does their coverage: the updates may concern individual titles or groups of titles, and they may be specific to a subscribed collection or include any new content from the publisher in question. These issues are pervasive in monographic e-series and e-book collections generally. Because of this variation, there can be no single method for monitoring new content availability that fits all situations.

Within the Libraries there are many parties who may be involved in dealing with the new content: selectors, catalogers, acquisitions and e-resources librarians, and Enterprise Technology. At present responsibility for monitoring new content, and for initiating the workflow processes that it should trigger, is not clearly allocated.

A special case of the notification problem is when a collection changes from print to electronic format. In these cases expectations that cataloging will continue uninterrupted regardless of format are not necessarily warranted. A switch of format can also raise other collection and access questions, such as whether publication will continue in parallel formats, and whether older titles will be digitized retrospectively. These questions can complicate communications further.

4.3.2 Collection maintenance

Ongoing maintenance of e-analytics is an issue as it is with all online resources. Maintenance of existing records may be needed in the following circumstances:

- An aggregator drops a collection from a package
- A subscription is discontinued
- A provider changes its URLs; or
- A resource disappears entirely.

Besides the need to perform this maintenance, resources need to be monitored so that we know of the changes in the first place. URLs often change without warning. With e-journals, much of this work is managed through SFX and MARCIt, but analytic records that are part of an e-journal title cannot be managed through the same automated processes as the records for the e-journals themselves. For these resources URLs should be checked and updated periodically. ALEPH has a URL checking feature that reports whether the URL connects, if it is redirected, or if there is a “page not found” error. The reports on redirected links or URLs with error messages require manual review and follow-up. The URL checker report has only been run a couple of times for serials and never for monographs. Technical Services has never had the resources to perform the manual clean-up work. If we were to provide electronic analytic cataloging, finding a solution for ongoing URL maintenance for non-serial titles would have to become a much higher priority.

The question of how to deal with maintenance of access to online resources is obviously also closely related to the question of the Libraries’ role in digital preservation.

4.3.3 Tracking treatment decisions

Because there is such a wide range of cases to deal with, it is necessary to document treatment decisions and associated workflow information (such as how notifications are being handled). There are a number of places treatment decisions could be tracked: Verde, Aleph order or order log, holdings, and the authority record. While none of these is an ideal location for the information, Verde might be the best choice, as the electronic resources management product contains information on our other electronic resources. Consistency in where to locate local administrative data is necessary for streamlining the process.

4.3.4 Projects vs. regular work

Many publishers are retrospectively digitizing their older print publications and making them available through their sites. Many selectors want to make these resources available to the same level as new online resources. However, this creates a large amount of cataloging work for one title, especially when the print versions of these resources are already in the catalog. Unless there are records available for batch loading, the titles then move into the project level (more than 25 items). As more and more of these situations occur, this will cause an even larger increase in cataloging work.

Because of the lack of clear guidelines for processing e-analytics and an absence of commitment to staff this work, a substantial backlog exists and there is considerable uncertainty about if, how, and when it will be dealt with. The current list of requests has about 50 titles, comprising serials and monographs.

5. Recommendations

5.1. Criteria for analytic treatment

The decision whether to analyze a given collection depends on many factors, including the staff resources available. It is therefore not possible to give absolute criteria for providing analytics. However, it is possible to identify key factors that should bear on the decision.

Items warrant consideration for analytic treatment, regardless of whether they are free or paid, if they fall into one or more of the following categories:

- Individual titles within the series or collection are distinctive.
- Access is not duplicated in discovery environments likely to be used by the target audience.
- We have a high degree of confidence that the resource is stable.
- We have acquired perpetual access.
- We have made a commitment to preserving the digital object.
- Record sets are available and of acceptable quality.
- There is a demonstrable need to provide authoritative identification of specific digital objects and the persons, organizations, or collections that are associated with them.

A change from print to electronic distribution should trigger a review of cataloging treatment.

5.2. Sources of records

5.2.1 Brief records

Where items would benefit from title-level catalog access but generally do not otherwise meet the criteria for analytic treatment, consider generating brief records from available data sources for use in the local catalog. If this option is pursued, a broader discussion of quality standards for record loads should take place.

5.2.2 Vendor services

Work with YBP to receive notification and cataloging for new additions to electronic collections not already covered by existing approval plans or firm orders.

5.3. Workflow and communication

Our discussions have clearly shown the need to increase transparency in the e-analytic process. Meetings with Academic Programs departments revealed many gaps of understanding on both sides. Technical Services should look for opportunities to work with Academic Programs to clarify processes and expectations. A good preliminary exercise would be to develop workflow charts that include both Academic Programs and Technical Services responsibilities to ensure understanding of the entire process from start to finish, and to clarify responsibilities. We also recommend the following steps.

5.3.1 Augment the e-resource form

The form should include information from selectors about cataloging preferences.

- Was this series or collection previously acquired by the Libraries in print?
- Are there individual titles in this free e-resource that need to be analyzed (i.e. cataloged separately)? Please list which ones.
- If yes to individual titles, will there be new titles added to in the future that require cataloging (one-time or ongoing)?
- If on-going, what will be the method of notification of new titles (i.e. link to email announcement list or RSS feed)?

The form should also include Technical Services project guidelines. Requests exceeding 25 titles that cannot be batch loaded fall under the guidelines for special projects and should be prioritized through the Collections Council process. This includes currently pending requests.

Once the revised form is available, selectors who have previously submitted requests should be notified so that they can provide the required information. Some of the requested information can and should be tracked in Verde, particularly regarding the availability of MARC records.

5.3.2 Document requests and their status

Explore options and select a solution for documenting cataloging requests and their progress. The chosen solution should be accessible to both selectors and Technical Services staff. There are two obvious possibilities: recording this information in Verde (the preferred option if workable), or creating and maintaining a spreadsheet to track requests for analytic treatment. At minimum, the following details should be recorded: collection title, number of titles in the collection, notification method for new titles, request status, selector name, and Technical Services contact.

5.3.3 Notification and monitoring

Establish a single email address to sign up for content alerts and newsletters from content providers. The email box should have a distributed monitoring system established, similar to the process used by Electronic and Print Serials in Technical Services.

For ongoing collections, identify the method of notification or monitoring that will be used to learn about new titles. Possibilities include:

- Publisher email alerts (i.e. to a Technical Services email account)
- RSS feeds from publisher or vendor sites
- RSS feeds from screen scraping tools (feedity.com, yahoo pipes, Yahoo Query Language)
- Gobi Alerts for new titles in series to which we have an ongoing subscription

It should be noted that new title and new record notifications are a major concern not only for e-analytics, but for ongoing record loads in general. Collaboration with ET and record loading staff to agree on a consistent means of notification is strongly recommended.

In each case responsibility should be assigned for monitoring new content. For selectively acquired items that task should fall to selectors; for collections Technical Services staff can play a role.

5.4. Timely decisions on cataloging requests

Technical Services and Collections Council should review all requests currently pending, assign each of them a status, and develop a plan to begin cataloging titles that are identified as high priority. New requests should also be reviewed and assigned a priority on a timely basis. Requests meeting the criterion for project work, i.e. exceeding 25 titles, should be evaluated as project requests through the existing Collections Council list.

6. Conclusion and Next Steps

E-analytics present not one kind of problem but rather a diverse range of problems. There is no one size fits all solution for e-analytics, and it is not realistic to aim to solve all of the problems at once. However, it is possible to assess priorities and look for quick wins. Here are some

proposals.

Achievable short-term goals include the following:

- Working with Technical Services Management, implement a pilot project to determine best practices (Appendix 5).
- Review currently pending projects for action
- Revise the electronic resources cataloging request form
- Determine the feasibility of documenting and reporting e-analytic cataloging requests in Verde. If Verde cannot be used for this purpose, create a spreadsheet to document cataloging requests for e-analytics.

Medium-term goals include:

- Investigate alerting and cataloging services from YBP
- Model and establish Academic Programs/Technical Services workflows for key cases, including signing up for alerts and allocating responsibility for monitoring them
- Identify appropriate uses for brief records and ways of generating them

Finally, we note that progress on some issues depends on developments in several related areas which are out of scope for this report but should be followed closely:

- Archiving of digital objects
- URL checking
- New discovery environments, e.g. WorldCat Local, Primo Central, Summon
- Identification of digital surrogates of print items and collections
- Archiving of digital objects

Appendix 1: Example Case Studies

Case study 1: Licensed Series analyzed in print, now moves online

1. *Special Publications, Geological Survey London* <http://sp.lyellcollection.org/current.dtl>

Overview: This monographic series published by a tier 1 geoscience society was moved online-only in 2010. The books are added periodically throughout the year for a flat fee that this library pays as part of the Lyell Collection subscription. The print titles had been analyzed, however, the electronic editions (of which are also being retrospectively added) are not being added to the catalog.

Needs: The ongoing notification and cataloging of e-only books published in this series since 2010. Also, the 300 or more older editions of the books that the library carried in print.

Assets: The publisher offers TOC email alerts of new books to any email address. Also, the publisher has send an excel file of the 300 e-books with URLs and titles. YBP slips exist for the individual titles in this series.

2. *Geological Society of America Monographic Book Series (3): Special Papers*, <http://specialpapers.gsapubs.org/>, *Memoirs*, <http://memoirs.gsapubs.org/>, **Field Guides**, <http://fieldguides.gsapubs.org/>

Overview: Monographic book series is published as online and print. Each book is a unique and distinct title, and the series editions come at irregular intervals. Online titles are paid as at a package rate. Individual print titles cost an additional 25\$ per title.

Needs: The selector would like to go online only; however, if the titles would no longer be individually analyzed, must opt to retain the print for discovery and access.

Assets: The publisher provides e-mail alerts and is also willing to send an XML list of new publications.

Case study 2: Series not analyzed in print then moves online (selected vs. ongoing)

1. *Water-Resources Data, Minnesota*
<http://www.dnr.state.mn.us/publications/waters/index.html>

Overview: Series received in print was a received annually since 1981 (Sys. no. 002326127). Moves online in 2006, is a separate pdf for each issue. This is a USGS deposit item.

Needs: Link to serial online would go to publications page and give directions to find water Year Data link on page. This must be done for all states or via a national site aggregator.

2. *Water Resource Investigations Report, Minnesota*

Overview: Series cataloged by title after 1983 (Sys. no. 000111494). Minnesota DNR discontinues the print version of the annual water resource report in 2007. Their web version is a pdf posted to an unstable website.

Needs: Archive our own copy of the DNR publications for long term access. Or work with third part (met with MDL to archive them, but they were not interested in born-digital, yet, only wanted to scan things not currently available online.)

Assets: The UDC. HathiTrust. A dedication/obligation to collecting Minnesota research.

Case study 3: Series/e-book package available online

1. *Springer E-books Package*

Overview: E-book package purchased by consortia for Springer published books that may or may not be duplicated as print holdings.

Needs: Seamless access to both print and electronic version of item from the catalog. Periodic updates.

Assets: Batch loading of e-book records, one time load of titles cataloged separately from print.

2. *Lecture Notes in Computer Science*

Overview: Monographic serial analyzed in print that is available in print+online

Needs: To switch acquisition to e-only

Assets: Records are loaded as part of the routine Springer E-book record loads.

3. *Clinics of North America*

Overview: BioMed stopped analyzing print series because users can find through PubMed index.

Needs: User base relies on PubMed already so titles in catalog not needed to duplicate effort for cataloging - get records as well.

Assets: PubMed, freed-up time.

4. *SourceOECD*, <http://titania.sourceoecd.org/vl=1463462/cl=22/nw=1/rpsv/home.htm>

Overview: Some but not all titles have records in OCLC. SourceOECD is being replaced by OECD iLibrary. Northwestern created an automated process for crawling the OECD site for titles, then searching OCLC for cataloguing records, or creating them where none were found. Around 1800 titles were cataloged in this way.

Assets: Northwestern has distributed the records at <http://www.library.northwestern.edu/public/OECD/>. Gary Strawn did the programming. However, this was a one-time project and the record set has not been updated since January 2009. It appears that updates in Northwestern's own catalog are being done manually. Northwestern did a similar project, also one-time, for the Annals of the New York Academy of Sciences.

- a. Northwestern OPAC view of SourceOECD - titles: <http://tinyurl.com/3y5hbqb>
- b. Northwestern OPAC view of ANYAS - titles: <http://tinyurl.com/2v12mxs>

Email alerts available, but not clear if they are specifically for *SourceOECD*. Presumably it would be possible to set up an email filter if *only* *SourceOECD* publications were

desired.

May be possible to do Feedity RSS: <http://feedity.com/rss.aspx/sourceoecd-org/UVBRVfpQ>

Case Study 4: Series received electronically though aggregator

1. AIP Conference Proceedings

Overview: Library does not subscribe to the periodically released conference proceedings (\$60,000/year) but does purchase print version on a title-by-title bases as related to ongoing research. The AIP Proceedings [0094-243X] are made available to us after one year embargo through a separate agreement with EBSCO as licensed through MINITEX.

Needs: These are high demand titles and due to their cost, very few can be purchased in print (~300/issue). Making the aggregated content available in the catalog would benefit users.

Assets: Because we do not purchase the title, don't get the analyzed cataloging through YBP. Access is not considered permanent because through aggregator and not direct purchase from AIP. MARCIt records provide series level cataloging. Problem here because uncertainty of ongoing access makes these weak candidates for cataloging unless there is an assurance that they will be acquired through other sources if access through EBSCO is lost. (i.e. prior collection management issue.)

Case Study 5: Free from source, online

1. Natural Resource Reports (NRR), <http://nature.nps.gov/publications/NRPM/nrr.cfm>

Overview: Print ceased and now online. Sporadic electronic reports added to website. Some bad links.

Needs: Unsure which titles to select and catalog. Stability of the URLs or that the agency will continue to keep on their website.

Assets: No alert. Not all titles listed are available online. RSS feed set up using feedity.com: <http://feedity.com/rss.aspx/nps-gov/UVBTUIJW>

2. Minnesota Geological Survey Reports and Maps

Overview: All publications by the University of Minnesota affiliated MGS have been scanned and loaded into the UDC.

Needs: These items have print records in the catalog, but the online editions not to be linked.

Assets: The UDC pdf is stable but the record is incomplete (i.e. bad). We could duplicate the print record and create new electronic version pointing to UDC. The MDL also has an electronic copy of the 600+ maps and they have pushed their content into Worldcat. copy. The electronic pdf is not stable however.

3. *Smithsonian Contributions to Botany*,
<http://www.sil.si.edu/smithsoniancontributions/Botany/>

Overview: Gov depository item (V., #, Yr. title, author. SI 1.29:). Current policy is based on GPO treatment. Not analyzed, GPO does not analyze.

Needs: This example would be a candidate for exception, since each title is distinctive and unique. Stability of the URLs in question. Not sure if that agency will continue to keep on their website.

Assets: All titles to date are in Smithsonian library catalogue (SIRIS) and have OCLC numbers. No alert available. URL follows pattern
http://www.sil.si.edu/smithsoniancontributions/Botany/sc_RecordSingle.cfm?filename=sctb-0093

4. *National Academies Press*, <http://www.nap.edu/>

Overview: The NAP publishes free books in pdf form to their website, print is available for purchase. These are of broad interest and high quality. Publisher, not a series.

Needs: Rather purchase print sporadically, a batch load of all NAP titles would be desirable.

Assets: Updates available by email, RSS, Twitter, and Facebook. Rights for distribution available for purchase - <http://www.nap.edu/permissions.html>.

5. *HathiTrust Monographs, not held in Print*

Overview: Scanned books from the Google project are available freely online (pending copyright decisions). Items already accessible from the catalog if held in print. Those titles not in print will not be visible.

Needs: users should not have to search in multiple places to get access to content that the HathiTrust partners have access to.

Assets: HathiTrust API, records available for loading, recommendations from MNCAT/HathiTrust linking group (2010):
https://netfiles.umn.edu/ul/Projects/ActiveProjects/MNCATHathi/Report_Record_Loads_Final.docx

Case Study 6: Archived born digital

1. *Technical Bulletin, Dept. of Natural Res., WI*

Overview: Born digital series but University of Wisconsin is archiving. Similar situation where we have the potential of doing something similar with UDC. Corresponding MN example above.

Needs: Archived version held in stable form/url. Notification system.

Assets: All titles available online from University of Wisconsin Digital Collections: <http://digital.library.wisc.edu/1711.dl/EcoNatRes.DNRBull>

2. *Wally Broecker's E-Books*

Overview: Faculty member recommends purchasing 4 e-books of famous environmentalist, but small university press only offers the pdf version via email. No electronic mechanism was available, so library purchased a CD version and faculty member keeps digital copy on computer to email to interested students. Titles are Greenhouse Puzzles (1998), Glacial World (2002), The Role of the Ocean, How to Build a Habitable Planet.

Needs: A stable way to make the content available to students electronically without checking out a CD.

Assets: E-reserves, UDC, web site.

Appendix 2: Meetings with Academic Program Departments

During the spring/summer of 2010, members of the group met with the Academic Programs directors and subsequently with the Academic Programs groups to discuss the issues around e-analytics related to their particular subject areas. Before the meetings a handout was sent out for review of the issues. Each meeting started with a brief overview of the charge and the discussions to date. Each Academic Programs group was then asked the following questions:

- What quantity of items are we dealing with? Do you have examples of items that you are dealing with? Why are you asking for cataloging (i.e. key access point, indexed in x databases, asked for by faculty, etc.)?
- How do you see patrons and staff accessing these items? How important is it to do this cataloging?
- The group is developing recommendations on policies around this issue. After reading the overview document do you have other issues or ideas that should be addressed?
- Anybody interested in a one-on-one meeting to discuss these issues?

Several reoccurring themes emerged from these meetings which are outlined below. There were similar themes to what the group had been discussing.

- The issues of e-analytics affects all groups, although some more than others. Each group provided additional examples but the concentration of need centered on government

information resources and the sciences. The type of resource also factored into this area. Some groups dealt more with e-book packages, some with a list of titles on a series website, still others had access only through a search database.

- There are differing ideas of what should be analyzed or not, even within the same Academic Programs group. These variances center on how to provide the best access to a title, and what is the best source. Several times the topic of whether patrons search our catalog for analyzed items or resort to resources such as Google Scholar were discussed. Other issues to be considered are how hard the item is to find, and is more than one person looking for this item.
- A question that came up with several groups was does it make sense to catalog an item on a website that might disappear in a couple months or the URL might change (free resources from the government or associations, for example). This question led to discussions about the Born Digital Group (Collections Council) and what online information could be saved locally.
- Some selectors felt they didn't know what questions to ask when it came to providing records or information to tech services about a new electronic resource they wanted analyzed. Where before the print volume dictated how the title was treated, the online versions can be handled in a variety of ways depending on the publisher's site, what information the selector wants to link to, and how this information correlates to standard cataloging processes.
- Incorrect URLs in the catalog is a large issue for both staff and patrons. As new resources are added, how does maintenance of the records, including correcting of URLs occur?

Appendix 3: Background

Facing budget reductions in 2003, the Libraries put in place a number of cost-savings measures, one of which was the reduction of analytic cataloging for print serials. Prior to implementing the cost-saving measure, Sue Zuriff and Bernie Karon (both members of Technical Services management at the time) wrote a report for Technical Services management identifying three types of analytics:

True monographic series: the monographic author and title are more prominent than the series title. These are books, not compilations of articles. Examples of monographic series include *Studien zur deutschen Literature* and *Lecture notes in computer science*.

Quasi-monographic, journal-like series: Items published annually (usually) with the series title more prominent than the monographic title. They are more frequently compilations of articles than true monographs, but with substantial content, a named author or editor, and an ISBN. Two examples are *Austrian studies* and *Wildlife* monographs.

Supplements to journals: These are either numbered in the series or issued with a numbered issue and using that number (for example, *Journal of paleontology*, v. 54, no. 4 supplement, "Paleoecological evaluation of late Eocene biostratigraphic zonation of the Pacific coast of North America").

Analytics were stopped for serial titles that met one or more of the following criteria:

- the serial is indexed and analytic-level metadata found in an abstracting and indexing database (i.e., EBSCOhost Academic Search Premier, Applied Science and Technology Abstracts, AGRICOLA, and ISI Web of Knowledge),
- the series title is more prominent than the monographic title or the monographic title is more generic than distinctive, or
- the circulation over the past ten years does not justify continuing analysis of the title.

Since 2003, other streamlining of Libraries workflows (S2A and S2A2) has moved print monographic series toward acquisition methods that include catalog records. Less intervention by Technical Services is necessary for monographic series that are ordered, received, and arrived at the Libraries “shelf-ready,” thereby reducing the cost of processing and freeing staff to take on other work. Another S2A/S2A2 process improvement suggestion that was accepted by the Libraries is the preference for electronic resources over print.

Then in 2007, members from the Cataloging Coordination Group (CCG) drafted a discussion paper titled "Cataloging Issues for Electronic Analytics" in response a number of requests from selectors to provide electronic analytic cataloging of electronic serials. It concludes that as the Libraries increasingly move toward electronic collections, and providing analytic cataloging for electronic series/serials is increasing prevalent/desired, there are still a number of issues that would need to be resolved before beginning manual analytic cataloging. Providing analytic cataloging for electronic resources would be a labor-intensive endeavor and the Libraries would need to weigh the cost in resources vs. the benefit and make decisions about where it fits in priorities.

Appendix 4: E-resources Form Questions for selectors and vendors

TITLE-LEVEL CATALOGING INFORMATION FOR E-BOOK COLLECTIONS

Collection Title:

Number of titles in collection:

Vendor:

- Are MARC records available for loading?
- If yes, is there an additional cost for MARC records?
- How can they be accessed?
- Name of vendor contact for records?
- Is a sample set available?
- Is customization available, and if so, what is the cost?
- Coordinate campuses that also subscribe?
- Is collection complete or are titles added on an ongoing basis?
- If yes, how often?
- Will titles be removed from the collection at any point?
- Does the vendor have a notification system for collection updates/changes?

- Who should we notify when loading or cataloging is completed?

Appendix 5: Pilot project

A lot of information is provided in this report regarding the current situation with e-analytics cataloging at the University of Minnesota. To ensure a greater understanding of how the issues identified above would mean in practical application within Technical Services, the E-analytics Working Group recommends a pilot project be implemented through Technical Services Management. Below is a general outline of potential parts of the pilot.

1. The following titles, and any others identified as suitable pilot project titles, should be analyzed.
 - Memoir (Geological Society of America : Online) (bib id: 5498885)
 - Field guide (Geological Society of America : Online) (bib id: 5597674)
 - Reviews in engineering geology (Online) (bib id: 5498889)
2. Information should be collected to determine time and resources necessary to analyze these titles.
3. Once the information is collected, it should be extrapolated to determine the level or staff and resources needed to manage a larger scale implementation of e-analytics cataloging.
4. This information should be provided to the Cataloging Coordination Group.