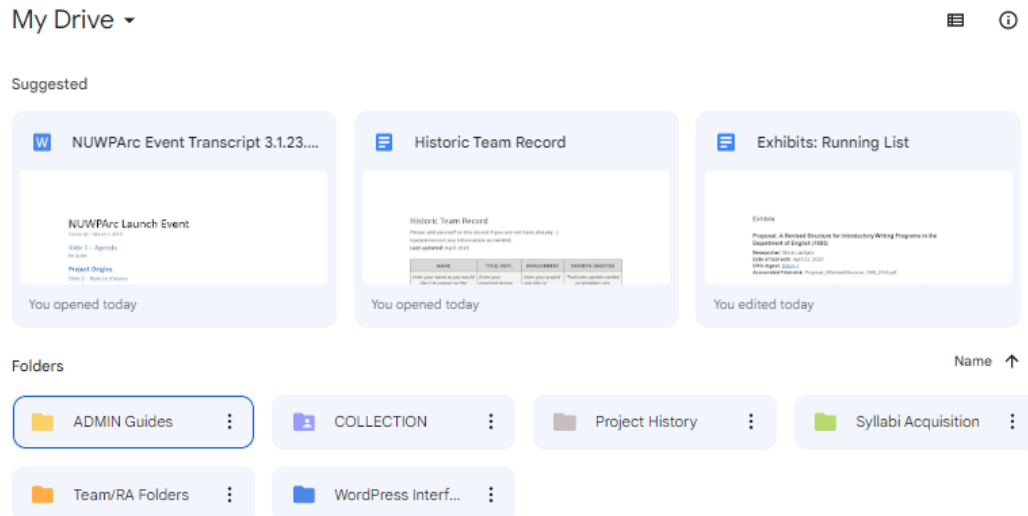


2022-2023 Ingest Notes

Kyle Oddis

[These procedures were advised by Sarah Sweeney & Drew Facklam in Northeastern University's Digital Scholarship Group. For reference, refer to Legacy records & Project Work Summaries documentation to observe changes in processes & procedures since 2020.]

[These notes are the "raw" version of discussions with the DSG; these have been expanded, refined, and included in the back-end G-Drive documentation to create a complete Administrator's Manual for future WPAs to use in managing NUWPArc following Kyle's departure from the PM role.]



DRS General Notes

- After project launch, revisit metadata end of April/May, since a lot has changed in the DRS
- We can have a more nuanced set of permissions
- Collection being public impacts how we upload the files, so need to separate permissions for different document types in public records ingests
- Separate out any student writing on back-end, so we can alert private permissions (for future)
- Separate out file formats when sending batches (word doc, pdf, .txt) since they have different intentions as objects; we can associate related files and connect after
- DRS considers PDF to be the preservation file format; redaction should carry over in the PDF that is generated – ingesting Word file will by default create a PDF
- Feature in DRS called "sets" – shopping cart kind of option – DRS in top corner, drop down menu, can create new sets and manage old sets, can browse, and there are buttons to add individual files to sets, can search and add files to sets

Creating Sets

- Sets can have distributed permissions; a group can have edit permissions for the sets you create
- Sets are dynamic and can change over time to add/remove to set; we can also make a collection record, zip them, and create one record that says this is a zip file for this collection, and that is an unchanging file that can't be added to/edited – but if I want a static teaching tool or a file for another specific purpose, that is something the DRS can hand create
- Put entry in spreadsheet with its own description or in a separate tab to describe in same columns, include a list of files that would go in that zip file
- There is a limit to zip file size, under 2 gigabytes, test it by zipping on my computer by clicking compress and seeing what the size is
- For batch uploads, we can submit files of any size; DRS can compress if needed and upload zip files of any size

Changes to Ingest spreadsheet template

Original

| | A | B | C | D | E | F | G |
|----|------------|--|---|-----------------|----------------------|--------------------|---------------------|
| 53 | | Advanced Writing in the Disciplines 309: Syllabus_Disciplines309_NoDate_0006.pdf | | Syllabus | James Weiss | | Syllabus, Course |
| 54 | 2006 | Eng U309.2: Advanced Writing in the Hu Syllabus_U309.2_2006_0007.pdf | | Syllabus | Christopher Weinmann | | Syllabus, Course |
| 55 | | English U307: Writing for Careers in the I Syllabus_U307_NoDate_0008.pdf | | Syllabus | Suzanne Richard | | Syllabus, Course |
| 56 | 2006 | English 307: Advanced Writing in the Sci Syllabus_English307_2006_0009.pdf | | Syllabus | Cecelia Musselman | | Syllabus, Course |
| 57 | | English U307: Advanced Writing in the S Syllabus_EnglishU307_NoDate_0010.pdf | | Syllabus | David Kellogg | | Syllabus, Course |
| 58 | 2006 | ENGU306 Advanced Writing in the Healt Syllabus_U306_2006_0011.pdf | | Syllabus | Alessandro Braidotti | | Syllabus, Course |
| 59 | 2006 | English U306 (Key #52417) Schedule Schedule_U306_2006_0012.pdf | | Schedule | | | Syllabus, Course |
| 60 | | ENG U306 Writing for the Health Profess Syllabus_U306_NoDate_0013.pdf | | Syllabus | Donna Scaglione | | Syllabus, Course |
| 61 | | English U306: Writing for Careers in the I Syllabus_U306_NoDate_0014.pdf | | Syllabus | Suzanne Richard | | Syllabus, Course |
| 62 | 2011 | Writing Center and Writing Instructor Cor Notes_WritingCenter_2011_0098.pdf | | Notes | | | Notes, Agenda |
| 63 | 10/19/2011 | Fwd: WPC agenda items...Alternatives to Email_WPCAgendaItems_2011_0099.pdf | | Email | Patricia Sullivan | Chris Gallagher | Email |
| 64 | 04/20/2011 | RE: Recommendations Letter_Recommendations_2011_0100.pdf | | Letter | ESL Inquiry Group | WPC | Letter |
| 65 | 04/20/2011 | RE: Teaching Network Spring 2011 Letter_TeachingNetwork_2011_0101.pdf | | Letter | Matt Noonan | Greg Zuch | Letter |
| 66 | | Writing Program Assessment Committee Proposal_WritingProgram_NoDate_0102.pdf | | Proposal | | | Outline, Proposal |
| 67 | | New Teaching Assistants and Instructors Guidelines_NewTeachingAssistants_NoDate_0103.pdf | | Guidelines | | | Guidelines, Writing |
| 68 | 06/02/2011 | Untitled Guidelines_Unitled_2011_0104.pdf | | Guidelines | | | Guidelines, Writing |
| 69 | 2006 | Writing Programs Workshop Guidelines_Workshop_2006_0105.pdf | | Guidelines | | | Guidelines, Writing |
| 70 | 2006 | important info for fall 06 Email_ImportantInfo_2006_0106.pdf | | Email | Kathleen Kelly | | Email |
| 71 | 2007 | The 2006-2007 Undergraduate Catalog Guidelines_2006-2007_2007_0107.pdf | | Guidelines | | | Email |
| 72 | 03/02/2006 | RE: desk copies for Northeastern Writing Email_DeskCopies_2006_0108.pdf | | Email | Patricia Sullivan | Brian Donnellan | Email |
| 73 | | The Department of English, the Writing Program Proposal_TheDepartment_NoDate_0109.pdf | | Proposal | | | Proposal |
| 74 | | Portfolios and Retrospective Assignment Guidelines_Portfolios_NoDate_0113.pdf | | Guidelines | | | Guidelines, Title |
| 75 | 01/05/2006 | First Day Assessment Procedures Memo_FirstDay_2006_0114.pdf | | Memo | Patricia Sullivan; | All instructors in | Memo, Guidelines |
| 76 | | End of Term Portfolio Review Guidelines_EndOfTermPortfolioReview_NoDate_0115.pdf | | Guidelines | | | Guidelines |
| 77 | 04/03/2006 | Handbook Webpage_Handbook_2006_0116.pdf | | Webpage | | | Webpage, Guide |
| 78 | | English 2/2A Writing Sample -- Instructor Guidelines_English2/2A_NoDate_0117.pdf | | Guidelines | | | Guidelines, Pron |
| 79 | 1980 | StudentWriting_Unitled_1980_0119.pdf | | Student Writing | | | Student Writing, |
| 80 | 1987 | Guidelines for Grading Papers Guidelines_For Grading Papers_1987_0120.pdf | | Guidelines | Ms. Shannon | | Guidelines, Eval |
| 81 | 1993 | Rubrics for Exit Assessment Guidelines_RubricsForExit_1993_0121.pdf | | Guidelines | | | Guidelines, Eval |
| 82 | | Packet of Readings for In-Class Entrance Guidelines_PacketofReadings_NoDate_0122.pdf | | Guidelines | | | Guidelines, Eval |
| 83 | | Writing Assessment: Terms and Definitions Guidelines_WritingAssessment_NoDate_0123.pdf | | Guidelines | | | Guidelines |
| 84 | 1988 | ENG 102: Course Outline (Winter Quarter) Syllabus_ENG102_1988_0124.pdf | | Syllabus | | | Syllabus |

Revised v.3

smartsheet

File Automation Forms Source Spreadsheet

| | INGEST STATUS | Date | Title | Filename | Filetype | Description | Category | Tags | Tagged for Interest | Author, institution-organization | Author, program-entity | First Author, personal-first | First Author, personal-last |
|-----|---------------|------------|--|--------------|----------|--|------------------|----------------------|---------------------|----------------------------------|------------------------|------------------------------|-----------------------------|
| 116 | Ingested | 2012 | The Writing Teachers' Sourcebook and Idea Book (Draft) | NJWPPar00117 | TXT | Contains activities, assignments, essay p | Handbooks | Course Design Curri | Yes | | | James | Stanfill |
| 117 | Ingested | 2005-08 | Teaching the New AWD | NJWPPar00118 | DOC | Contains information on AWD and offers | Handbooks | Course Design Curri | Yes | | | David | Kellogg |
| 118 | Ingested | 2005-08 | Teaching the New AWD | NJWPPar00119 | TXT | Contains information on AWD and offers | Handbooks | Course Design Curri | Yes | | | David | Kellogg |
| 119 | Ingested | 2012 | The AWD Toolkit, Online Edition | NJWPPar00120 | PDF | Offers an online edition of the AWD Tool | Handbooks | Curriculum Technolo | Yes | | | David | Kellogg |
| 120 | Ingested | 2012 | The AWD Toolkit, Online Edition | NJWPPar00121 | TXT | Offers an online edition of the AWD Tool | Handbooks | Curriculum Technolo | Yes | | | David | Kellogg |
| 121 | Ingested | 2006-04-03 | Handbook | NJWPPar00122 | PDF | Printed out webpages capturing the onli | Handbooks | Curriculum First-Yea | Yes | | | | |
| 122 | Ingested | | Assistant to the Writing Program | NJWPPar00123 | PDF | Describes the role of the Assistant to the | Job Descriptions | Labor Program Adm | Yes | | | | |
| 123 | Ingested | | Associate Coordinator Positions | NJWPPar00124 | PDF | Describes the roles of Associate Coordin | Job Descriptions | Labor Program Adm | Yes | | | | |
| 124 | Ingested | | GTA Positions Call (Draft) | NJWPPar00125 | PDF | Describes the roles of Assistant to the W | Job Descriptions | Graduate Students i | Yes | | | | |
| 125 | Ingested | | Faculty Director of Writing Center Position | NJWPPar00126 | PDF | Describes the role of Faculty Director of | Job Descriptions | Labor Program Adm | Yes | | | | |
| 126 | Ingested | | Writing Program Overview | NJWPPar00127 | PDF | Describes the duties of Writing Program | Job Descriptions | Labor Program Adm | Yes | | | | |
| 127 | Ingested | 2011 | Student Learning Goals (Draft) | NJWPPar00128 | PDF | Draft with notes on eleven student learn | Learning Goals | Program Development | Yes | | | | |
| 128 | Ingested | 2012 | Student Learning Goals | NJWPPar00129 | PDF | Eleven student learning goals for the Wri | Learning Goals | Program Development | Yes | | | | |
| 129 | Ingested | 1997-11-13 | Writing Programs Committee Meeting | NJWPPar00130 | PDF | Minutes cover class visits/update, subcor | Meeting Minutes | Program Administrat | Yes | | | | |
| 130 | Ingested | 1999-10-25 | Writing Programs Committee Meeting | NJWPPar00131 | PDF | Minutes cover class visits for first-year Ti | Meeting Minutes | Program Administrat | Yes | | | | |
| 131 | Ingested | 1999-03-04 | Writing Programs Committee Meeting | NJWPPar00132 | PDF | Minutes cover section management, funi | Meeting Minutes | Program Administrat | Yes | | | | |
| 132 | Ingested | 2000-03-14 | Writing Programs Committee Meeting | NJWPPar00133 | PDF | Minutes cover Spring Quarter courses/c | Meeting Minutes | Program Administrat | Yes | | | | |
| 133 | Ingested | 2000-04-06 | Writing Programs Committee Meeting | NJWPPar00134 | PDF | Minutes cover Spring Quarter opening, c | Meeting Minutes | Program Administrat | Yes | | | | |
| 134 | Ingested | 2000-04-27 | Writing Programs Committee Meeting | NJWPPar00135 | PDF | Minutes cover class visits, Topic for Sprin | Meeting Minutes | Middle-Year Writing | Yes | | | | |
| 135 | Ingested | 2000-01-31 | Writing Programs Committee Meeting | NJWPPar00136 | PDF | Minutes cover Winter Quarter opening, c | Meeting Minutes | Middle-Year Writing | Yes | | | | |
| 136 | Ingested | 2004-02-19 | Writing Programs Committee Meeting | NJWPPar00137 | PDF | Minutes cover rescheduling, class visits, a | Meeting Minutes | Enrollment Program | Yes | | | | |
| 137 | Ingested | 2008-10-30 | Writing Programs Committee Meeting | NJWPPar00138 | PDF | Minutes cover class visits, Peer Tutors, a | Meeting Minutes | Program Administrat | Yes | | | | |
| 138 | Ingested | 2004-10-27 | Writing Programs Committee Meeting | NJWPPar00139 | PDF | Minutes cover a new Interdisciplinary All | Meeting Minutes | Program Administrat | Yes | | | | |
| 139 | Ingested | 2008-09-08 | Writing Programs Committee Meeting | NJWPPar00140 | PDF | Minutes cover scheduling, goals for the | Meeting Minutes | Program Administrat | Yes | | | | |
| 140 | Ingested | 2010-03-24 | Writing Programs Committee Meeting | NJWPPar00141 | PDF | Minutes cover updates on Placement sitc | Meeting Minutes | Graduate Students i | Yes | | | | |
| 141 | Ingested | 2010-03-21 | Writing Programs Committee Meeting | NJWPPar00142 | PDF | Minutes cover updates on change, Gradfr | Meeting Minutes | First-Year Writing | Yes | | | | |
| 142 | Ingested | 1993-03-15 | Writing Programs Committee Meeting | NJWPPar00143 | PDF | Minutes cover classroom visits, survey of | Meeting Minutes | Middle-Year Writing | Yes | | | | |
| 143 | Ingested | 1994-06 | Writing Programs Committee Meeting | NJWPPar00144 | PDF | Minutes cover Writing Committee Chair i | Meeting Minutes | Program Administrat | Yes | | | | |
| 144 | Ingested | 2004-01-14 | Writing Programs Committee Meeting | NJWPPar00145 | PDF | Minutes cover AI/ID planning and positio | Meeting Minutes | Enrollment Program | Yes | | | | |
| 145 | Ingested | 1993-01-05 | Writing Programs Committee Meeting | NJWPPar00146 | PDF | Minutes cover class visits/doctoral candi | Meeting Minutes | Program Administrat | Yes | | | | |

- Standardized dates: no padding
 - Go YEAR/MONTH/DAY, most general to most specific, dashes are better than slashes – that would save a ton a post editing time; if no dates, leave blank in spreadsheet field
 - Month and day are not required; month is required if you have a day
- Title: sufficiently descriptive, but do I want the colon to truly be what is different between the title and subtitle – how am I using the colon every time?
- Capitalization: standardize “Re” vs “re” – soft recommendation as opposed to dates
- Do I want the file name visible in the record or not?
 - Do I want the .pdf there, or would I use that as a unique alphanumeric string that would identify that file?
- Turn document types (genres/categories) into pluralizations
- Author name format: reduce variability
 - Separate out first and middle name from last name, and salutations from all of those, so a name would have 3 diff fields, first and middle column, last name column, terms of address column (e.g., Ms. Jr. Sr. etc.)
 - Example: R.B. Gill – to imply R.B. is first and middle and Gill is last; separate fields
 - There are corporate names, so if we were re-writing metadata, this would be suggestions, not as much work as going through the dates
 - Columns for first middle, last, designation, and then a separate column for corporate names (depends on how many names there are)
 - Faster turnaround!
- In content tags, eliminate overlap between tags and document types
 - Genres – what it IS vs what it’s about
 - Document tag is genre, but content tag is not subject
 - Remove genre and replace what it is with what it’s about for content tags
 - Ex. Row 3, chart is in doc type, but all 3 can be present in genres; 100% necessary needs at least one content tag

- We can apply a collection wide subject to make sure they all have subject; searching DRS wide makes collection more visible
 - We can have as many doc types as we want, but controlled list of subject tags
 - Think about it as an iterative process
- **Metadata practices have changed substantially**

Early on perspective was “no one wants to talk about structuring data in a useful way, so give it to DRS and we’ll figure it out” HOWEVER— NUWPArc focus is thinking about composing data structure from initial phases! Woo! *Important for future PMS!!!*

Clarifications

1. Each author needs their own set of fields. There technically isn’t a maximum, but I usually limit the creators to six names.
2. Any audio or video files also need an associated image file to act as a thumbnail for it in order to be ingested into the DRS. For example, you can see in this [collection](#), that all the files have the same initial image for the audio recordings.
3. The file type doesn’t matter for ingest, so you can send them separately, if that’s how you already have it organized, or you can send it all together in one spreadsheet—just make sure the filename has an **exact match** in a column in your spreadsheet, including the file extension. The easiest way for us to access the files would be via a Google Drive/DropBox/Teams folder, sending them via email would have a lot of file size issues.
4. If you want files deleted, because you have new versions of the files, let DRS know and send along a spreadsheet with the permanent URLs.
5. We can add to existing metadata at a later date, BUT must meet to discuss how to best format spreadsheets to make sure that we have a valid match point to direct the overlay.

7 to 10-day delay between sending spreadsheets and when changes are available in the DRS.

Overlays

- Creating new data in spreadsheets is where excel shines, but OpenRefine is powerful for doing massive changes for columns and generating new columns—so great for editing existing spreadsheet and facilitating changing it.
- **Doesn’t matter if the file name is descriptive in the spreadsheet**– only thing that’s important is that it’s a match point in the drive.
 - Each row of the spreadsheet is a separate record; the file name column will generate the name of the XML record – the only thing that matters is the filename listed in the cell is an exact match.
- File naming conventions are important, but if there’s not a reason to maintain them, that’s not going to affect DRS process
 - Don’t ever make a filename something that’s changeable – like genres! That list might change/expand, etc. so that would create a lot of work if genres get more granular (yep)
- Filenames CHANGED to alphanumeric strings; more viable for genres
 - Don’t need to have that info in file name to do the same work twice

What does this mean for overlays?

- What matters for overlays is that there’s a match point
- The DRS is going to change file names to match DRS’ alphanumeric strings

- Can export a manifest of collection– that has original filename with the **PID** (match point), so if adding metadata, take that info and expand upon that with new metadata

Editing Existing Data or Adding Additional Data

Adding Metadata

- I would want to add additional metadata (like more tags), so just take the manifest and add the additional subjects, and that's it!
- Change column type to text in excel so I can remove the underscore in column c so it doesn't edit out the zeros – nothing wrong with having the underscore either
- Needs the file extension in the filename, so if I take out the extension in the title, just add a column with the extension even if the filename is just the alphanumeric string
- Same number with a different extension

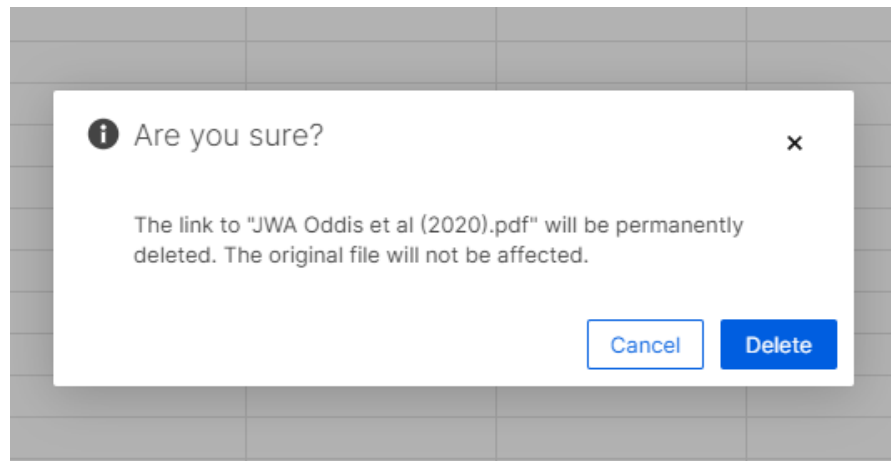
Existing Metadata Overlay

- Overlaying existing metadata can get slightly more complicated.
- Rules of having match point is the same, but if I do want to change existing metadata—say I want to change 30/500 titles, but I want to change all subjects in all 500– I can't have a spreadsheet with 30 titles changed and the rest blank
- I would need to include the rest with the original info and just the ones changed that I want changed — saves a lot of time, we're talking *hours vs days*
- After ingest, will send manifest that includes PIDs

Avoid overthinking filenames! Advice: Just make it a string we know we **never** have to change again :)

G-Drive Back-end

- NUWPArc 2023 filenames are all alphanumeric strings beginning with NUWPArc and a string of 5 numbers (e.g., NUWPArc00001)
 - That string is autogenerated in Smartsheet; I add the metadata to columns, then export the Smartsheet to a google doc or excel sheet
 - Smartsheet allows associating G-Drive files directly with lines in the sheet without affecting the G-Drive files so I can easily cross-reference files for tagging & sorting



- To ensure exact matches, I organized all original files (that were titled with descriptive filenames) into the new administrative G-Drive, which is organized by document type into folders that contain the records titled with their new alphanumeric strings

