

# IOWA STATE UNIVERSITY

## SPECIAL COLLECTIONS AND UNIVERSITY ARCHIVES

**Jump In 3 - Fall 2014 to Spring 2015**

### PROJECT BACKGROUND AND PREVIOUS PRACTICE

The Iowa State University Special Collections and University Archives (SCUA) has been acquiring digital materials on storage media for many years. Realizing that digital stewardship was becoming a crucial element of the archives program, the department successfully requested a new Digital Archivist position to develop and oversee digital stewardship efforts. I was hired in August of 2014 to fulfill this role.

Prior to my hire SCUA staff had been making do with what was available to us: network storage, two [DROBO](#) drives, and staff time created by borrowing from their other duties. We had begun setting aside digital "originals" in dedicated storage spaces and tracking digital accessions in a table within our collections management databases.

Accession Number	CD	DVD	USB Drive	Floppy - (3 1/2")	Floppy - (5 1/4")	Other	Description of Files	Label Info	Date of Accession	Transferred to Network Storage?	Location
2014-314	✗	✗	✗	✗	✗	✓	4 doc, 5docx, 1 xlsx, 8pdf Unknown. The files were not accessed. File size listed above is only an approximation		12/18/14	Yes	<a href="#">V:\Elect and Cor 2015\20</a>
2014-193	✗	✗	✗	✓	✗	✗	Unknown. The files were not accessed. File size listed above is only an approximation	administrative files, vitas	7/29/2014	No	
2014-050	✗	✗	✗	✗	✓	✗	Unknown. The files were not accessed. File size listed above is only an approximation	FIEI annual reports, program files	2/26/2014	No	
2014-142	✗	✓	✗	✗	✗	✗	Disc is a DVD-RW. Files were not accessed.		6/5/2014	No	
2014-325	✗	✗	✗	✗	✗	✓	Images from campus events taken by Bob Elbert, University Photographer, and other staff and images used for Inside Iowa State		12/24/2014	Yes	<a href="#">N:\Univ</a>

Figure 1 -The table we use to track the storage location of digital materials.

In late 2013 we began running hashes (mostly MD5) and acquiring directory prints via Karen's Directory Printer.<sup>1</sup> The results of these analyses were also saved in network storage. However, there were not fully established ingest procedures and no maintenance activities such as migration or checking fixity.

Most of our digital materials arrived on storage media as part of hybrid collections, but we are increasingly seeing small transfers and donations via email. In addition, we have received large transfers of image files from University Marketing via external hard drives which form the bulk (in GBs) of our recent digital acquisitions. These files are added to the DROBO drives and copies are made accessible via [Flickr](#).

**1**

<sup>1</sup> Karen Kenworthy developed several free and useful programs for digital preservation. She passed away in 2011, but there is currently a [volunteer effort](#) going on to re-create her website. In the meantime, her software can still be downloaded through several sources at your own risk. I make no claims as to the safety or legality of these sites.

With my arrival our plan has been to:

- 1) identify digital preservation and stewardship needs for both existing and future acquisitions
- 2) identify technical requirements for, and oversee the installation of, a digital preservation workstation
- 3) test and select digital preservation and digital stewardship software
- 4) create policies and procedures to shape a robust digital program

Participation in *Jump In 3* supports all of these functions as it helps us identify the amount and type of formats we previously acquired (1), provides information on what type of peripherals and software we need to perform full preservation (2, 3), and suggests policies and procedures for future acquisitions (4).

## SCOPE OF SURVEY

At SCUA we have two primary collection divisions. Our manuscript collections (MS) are collected materials that reflect the major research and education areas of Iowa State University: Life Sciences, Agriculture, and Engineering. Our university archives materials form the documentary heritage of the activities, experiences, and functions of Iowa State University and the campus community. These are typically comprised of materials transferred from university units or the personal papers of alumni and faculty.

Our university archives materials are larger and more complex than our manuscript materials, so I decided to test out the method I developed for this study by starting with manuscript collections, many of which are older and smaller than recent university archives materials. I assumed that the manuscript collections would yield a more manageable survey frame.

## PARTICIPANTS AND LOGISTICS

The survey was designed by myself and carried out by SCUA student worker Victoria Pohlen, a Senior double-majoring in Biology and History.

### TIMELINE

Our timeline was as follows:

December 2014 - February 2015 I constructed the method, the instructions, the documentation, and the survey binder.

Using a modified version of our manuscripts status table, I identified the location and format of finding aids for each MS and constructed a formula to count the available locations for each finding aid.

Collection Title	online?	in template?	electronic?	in binders?	FA current?	FA?
McCall Family	YES	YES	YES	YES	YES	5
Rae, George	YES	YES	YES	YES	YES	5
Hearst, Charles J.	YES	YES	YES	YES	YES	5
Hanyan Family Pa	YES	YES	YES	YES	YES	5
Shepard, Hugh H.	NO	NO	NO	YES	NO	1
Olsen, Nils Andre	YES	YES	YES	YES	NO	4

On another sheet of the Excel workbook I created a log for Victoria to record the locations of digital storage media for each collection. I transferred over the collection lists and created a formula that would tell her if a finding aid was available for that collection.

MS Number	Collection Title	FA?
MS-337	de Planque, E. Gail	Yes
MS-341	League of Women Voters of Ames	Yes
MS-342	International Textile and Apparel As:	Yes
MS-349	American Statistical Association	Yes
MS-350	Ames Town and Gown Chamber Mus	Yes
MS-413	Center for Rural Affairs	Yes
MS-420	American College of Veterinary Path	Yes

Victoria searched the finding aids for the following search terms: *disk, disc, CD, DVD, floppy, drive, USB, flash, "computer tape", "data tape", digital*. (See Appendix 1)

Victoria added the location (box/folder) of the storage media to the MS survey log:

MS Number	Collection Title	FA?	Date FA		FA Digital	digital loc
			checked			
MS-337	de Planque, E. Gail	Yes	3-Feb-15	Yes		52, 67, 72, 77
MS-341	League of Women Voters of Ames	Yes	3-Feb-15	Yes		43/16
MS-342	International Textile and Apparel As:	Yes	3-Feb-15	Yes		31/57
MS-349	American Statistical Association	Yes	3-Feb-15	Yes		193/24-25, 193/64, 194/15, 194/24, 196/1,
MS-350	Ames Town and Gown Chamber Mus	Yes	9-Feb-15	Yes		17/19, 18/23, 19/1, 22/12, 23/5, 25/11,
MS-413	Center for Rural Affairs	Yes	9-Feb-15	Yes		135/9, 144/2-4, 146/17
MS-420	American College of Veterinary Path	Yes	4-Feb-15	Yes		18/17, 20/39-42, 22/8, 22/45, 23/7-8, 28/1,
MS-422	Chrystal, John	Yes	4-Feb-15	Yes		13/8
MS-423	American Association of Feed Micros	Yes	9-Feb-15	Yes		4/37, 4/42

February 2015

I attempted to devise a formula for visually assessing the data stored on CD and DVD (see next section).

March 2015 - April 2015

Victoria physical scanned the identified folders and filled out a JIMS ("Jump In - Manuscripts") record for each item she found (See Appendix 2)

May 2015

Victoria graduated and I wrote this report.

## LOGISTICS

I devised a binder for Victoria to use while completing the physical survey. It has four sections:

- 1) Instructions (see Appendix 2)
- 2) A log for format counts by collection surveyed.
- 3) An entry form for each individual storage object.
- 4) A field guide to the media

Victoria traded off this task with other tasks for ergonomic reasons - standing, stooping, reaching, etc. are tiring or can lead to strain or injury if done for long stretches of time.

The physical survey went quickly (about 4 weeks) and was unproblematic. Victoria had a few questions (how to handle recordkeeping for large sets of conference proceedings on disk, what to do when she couldn't tell what kind of CD or what the capacity was, etc.), but for the most part she was able to proceed without much additional guidance. I

think it was helpful to have the same student work with me throughout the project so that she knew what she was looking for, why we were doing it, and what to expect with each box she opened. She started the survey with the materials we have on site and saved all of our off-site materials until the end of her survey.

## AN IDEA THAT WENT NOWHERE

I thought the numbers for GBs of content arising from this effort would be more compelling if they were precise rather than reliant on the (unlikely) possibility that each media storage object was full to its maximum capacity. The fullness of optical disks (CDs, DVDs) is discernable by the change in color of ink on the underside of the disk. Being that disks are circles, I hoped that I would be able to estimate the amount of MBs/GBs on each disk based on the area encompassed by the different color of ink. In other words - figure out the area of the circle making up the filled space of the disk and then use that area (as a fraction of the whole) to get a number for the actual storage used. See appendix 3 for an image of my attempt at devising a formula.

The problem is that although the data is encoded in a spiral at a consistent speed (the disk drive reads at a variable speed as the spiral grows wider), I did not account for track pitch (distance between tracks of the spiral) or for the space taken up on each frame by what is roughly equivalent to "metadata." In brief, the physical area of user data in the inner portion of the disk does not seem to be a constant with the physical area of the user data in the outer portion of the disk since this "metadata" may take up unknown area in any given physical part of the disk.

I do not have the math or engineering skills to pursue this further, but would love to hear from anyone who might be able to devise a formula.

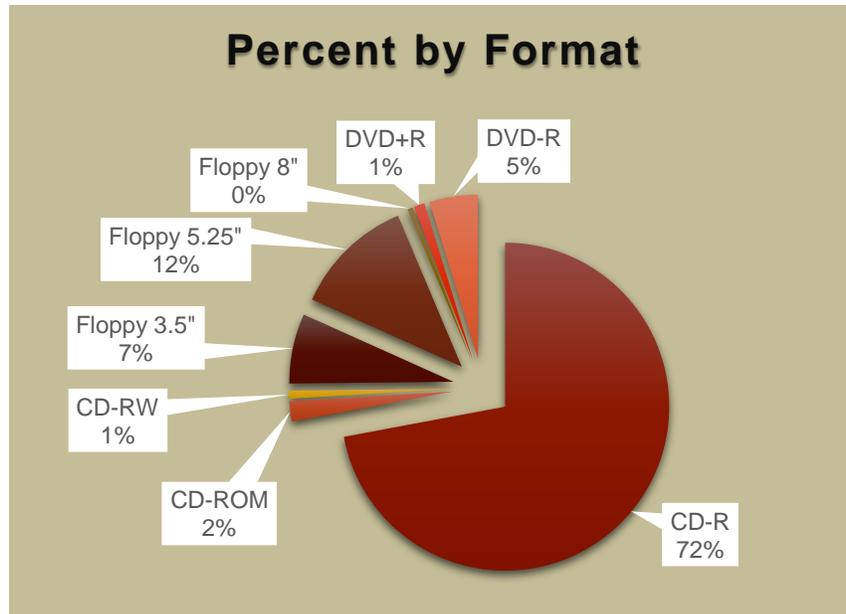
## OUTCOMES AND ANALYSIS

Although we have 653 manuscript collections, only 19% of them indicated digital materials in the finding aids. Victoria found 758 storage media objects within the 124 collections surveyed. Of these 124 collections, 25 ended up not having digital media objects within them. This could be for a variety of reasons: the finding aids might be out of date (i.e. materials had been reformatted, printed, or were returned to donor), physical separation sheets were missing (materials may have been stored in the vault or in a dedicated disk box in another area of the stacks), or the materials found were audiotapes rather than "data tapes." Other scenarios require additional investigation. Happily, the majority of materials were in good shape physically. A few had minor scratches or smudges. One item had a small crack.

I planned for us to discover the following types of storage media:

CD-R	DVD+R	Blu-Ray	USB flash drive
CD-ROM	DVD-R	miniBlu-Ray	Memory card
CD-RW	DVD+RW	Floppy 3.5"	Memory Stick
miniCD	DVD-RW	Floppy 5.25"	ext. hard drive
Zip disk	miniDVD	Floppy 8"	Other

We actually only discovered CDs, floppies, and DVDs. The least common formats discovered were a DVD+RW, a miniDVD, and a miniBlu-Ray. The most common formats were CD-R (544), 5.25" floppy (90), and 3.5" floppy (52).



We also discovered that most collections will only have a few storage media objects. The bulk of the objects surveyed were found in three collections: People United for Rural Education (PURE)(MS 300), Western Veterinary Conference (MS 305), and the Institute of Mathematical Statistics - Bulletin (MS 324). MS 300 had 71 5.25" floppies, MS 305 had 481 CD-Rs, and MS 324 had 30 3.5" floppies. The remaining 176 objects found were spread throughout 96 other collections.

Our findings were as follows:

Format	Count	Max MB	Potential storage needs
CD-R	544	700	380800
CD-ROM	15	700	10500
CD-RW	6	700	4200
miniBlu-Ray	1	7987.2	7987.2
Floppy 3.5"	52	2.88	149.76
Floppy 5.25"	90	1.2	108
Floppy 8"	4	6.2	24.8
DVD+R	8	4474.88	35799.04
DVD-R	36	4474.88	161095.68
DVD+RW	1	4474.88	4474.88
miniDVD	1	4812.8	4812.8
<i>Total storage media:</i>	<i>758</i>		
<i>Nothing found:</i>	<i>25</i>		
<i>Total potential storage:</i>		MB	609952.16
		GB	595.66
		TB	0.58

## CHALLENGES

This process was largely painless -- it worked as I designed it and Victoria reported no major problems.

Small hitches were:

- The finding aid search was not always accurate.
- It is unrealistic to distinguish sub-types of CDs and DVDs consistently.
- Many storage media, disks in particular, have no distinguishing features.
- Plan for the ergonomic aspects of the work.

## FUTURE STEPS

This was a very useful exercise that generated empirical data that can be used to support equipment and storage requests, the establishment of iterative goals, and software selection and adoption.

I plan to follow-up this activity with the following projects:

- An identical survey of our university archives materials
- Ensure that the digital stewardship workstation is configured to address all found formats. It currently does not have capability for working with larger format floppies.
- Check the finding aids where no digital content was found to figure out why they came up as having digital materials in our initial search.

Submitted on May 6, 2015 by

Kimberly Anderson, PhD  
Digital Archivist

Special Collections  
Iowa State University  
Ph: 515-294-8590  
Fax: 515-294-3186  
kda@iastate.edu

main: <http://www.add.lib.iastate.edu/spcl/index.html>  
blog: <http://isuspecialcollections.wordpress.com/>

*With immense thanks to Victoria Pohlen for her excellent work on this project.*

# APPENDIX 1

## Jump in 2015: Manuscripts Finding Aid Survey

### INSTRUCTIONS

#### Goal:

We will be participating in the Society of American Archivist's 2015 "[Jump In](#)" initiative. We'll be inventorying the digital materials we have on storage media located throughout our manuscripts collections. In other words - we're going to find all the CD-ROMS, hard drives, floppy disks, USB sticks, etc. that we've got in our manuscripts collections!

#### Your Task:

**Check the finding aids to see if there are any digital storage media.**

We're starting off by identifying which manuscript collections have digital materials. A spreadsheet has been created which shows whether or not a particular manuscript collection has a finding aid in some format. If a finding aid does not exist, the cell will indicate "**No**":

MS Numb	Collection Title	FA?	Date FA checked	FA Digital?	digital loc
MS-016	Citizens Committee for a Constitutional Convention (Iowa)	Yes	11-Nov-14	No	
MS-017	Gaudineer, Lee H.	Yes	11-Nov-14	No	
MS-018	Harbor, William H.	Yes	11-Nov-14	No	
MS-018A	Millen/Harbor, Representatives	No			
MS-019	Fort Dodge Laboratories	Yes	11-Nov-14	No	
MS-020	Polk County [Iowa] Equity Circuit Court Cases	Yes	excluded - pre 1950		
MS-021	Darrow, Wayne	Yes	11-Nov-14	No	
MS-022	Camp, John Bryant	Yes	11-Nov-14	No	
MS-023	Dancey, Jesse Samuel	Yes	11-Nov-14	No	
MS-024	Barnett, Clay C.	Yes	11-Nov-14	No	
MS-025	Nelson, Rev. Laurence N.	Yes	11-Nov-14	No	
MS-026	Iowa Conservation Education Council	Yes	11-Nov-14	No	
MS-027	Weed Science Society of America	Yes	11-Nov-14	Yes	34/17
MS-028	McNurlen, Keith A.	Yes	11-Nov-14	No	

You will need to check all the collections that *do* have finding aids.

#### Steps:

1. Open the spreadsheet:  
<V:\Projects\JumpInInitiative2015\MSSurvey30Jan2015.xlsx>
2. Find the next empty space in Column D (Date FA checked)
3. Check the corresponding row in column C (FA?).
4. If there is a finding aid, get the collection number (column B) and go to the "manuscripts" tab in the workbook.

C	D	E	F
FA?	Date FA checked	FA Digital?	digital loc
No			
Yes	11-Nov-14	No	
Yes	excluded - pre 1950		
Yes	11-Nov-14	No	
Yes	11-Nov-14	Yes	34/17
Yes	11-Nov-14	No	

- Check columns D, F, G in the "manuscripts" tab to see where the finding aids are located. In this example, you could check the online finding aid for MS-027, but would need to check the binder for MS-031:

B	C	D	E	F	G
MS Numb	Collection Title	online?	in template?	electronic?	in binders?
MS-027	Weed Science Society of America	YES	NO	YES	YES
MS-028	McNurlen, Keith A.	NO	NO	NO	YES
MS-029	Armstrong, Ruth Gallup	YES	YES	YES	YES
MS-030	Manuscript Leaves and Ephemera	YES	YES	YES	YES
MS-031	Spring Hill Coal Company	NO	NO	NO	YES
MS-032	Dorset, M. (Marion)	YES	YES	YES	YES
MS-033	Berry, Don L. (Don Lytton)	YES	YES	YES	YES
MS-034	Iowa-Yucatan Partners of the America	NO	NO	NO	YES

- You only need to check **ONE** of these finding aids. If there is a finding aid online, use that. If no finding aid is available online, then use the electronic version in the V://. Check the binders only if there are no other options.
- Make sure the collection dates start at 1950 or later. If the collection ends before 1950, then you can enter "excluded - pre 1950" in the "Data FA Checked" column. Go back to step 2.
- ctrl-F and look for these terms: disk, disc, CD, DVD, floppy, drive, USB, flash, "computer tape", "data tape", digital (you will be doing a total of 11 searches on each finding aid - 1 per term)
- If you find anything that is digital storage media, go back to the "MSLog" tab and enter "Yes" in column E (FA Digital?)
- Enter the box and folder number in column F (digital loc)
- If you need to enter multiple boxes and folders, you can add additional lines in the cell by entering Alt-Enter. Great! Start again at step 2.

## APPENDIX 2

### JUMP IN 3 INSTRUCTIONS

#### ***SURVEY OF COMPUTER STORAGE MEDIA IN MANUSCRIPT COLLECTIONS***

##### *GOAL OF PROJECT:*

To assess the quantity and kind of digital files we might have stored in the manuscript collections.

##### *HOW WE'LL MEET THE GOAL:*

We will find each storage media item located within our manuscript collections. We will identify the media type, manufacturer, max capacity of the object, its label markings, an e-record ID (if one was assigned), and any additional comments.

##### *STEPS:*

1. Check each box and folder listed in the "Format Counts " (yellow paper)
2. There should be at least one item in each box that qualifies. If you have found digital/computer storage media, then fill out an entry on the "shelf survey" (white paper) form **for each object**. Assign the next whole number as the JIMS # (jump in Manuscripts ID #) for each object.
  - a. There are two entries per page, and four entries per sheet.
  - b. Some collections have more than one object. You only need to write the brief title of the collection on the first entry for each MS.
3. Use the "Field Guide to Digital Storage Media" to get an idea of what you're looking at. These sheets will have rough dates, dimensions, and major manufacturers listed.
4. After you have finished entering all objects for a folder, fill out the appropriate line in the "format counts" form. You will note the total # of each format type found in that folder, along with any associated e-rec IDs. Make sure you transfer over the Jump In ID#s from the "shelf survey" sheet. Enter a JIMS# range if there is more than one item per folder.
5. Repeat until the format counts sheet is complete.

See next page for examples.

JUMP IN  
ID#:  
(JIMS)

31

MS#: 1000

Box#: 5

Folder #: 22

Collection Name/brief title: Cy Papers

Media Type	Comments	Max Capacity	MB/GB
CD-R <input type="checkbox"/> Blu-Ray <input checked="" type="checkbox"/> DVD+R <input type="checkbox"/> USB flash drive <input type="checkbox"/> CD-ROM <input type="checkbox"/> miniBlu-Ray <input type="checkbox"/> DVD-R <input type="checkbox"/> Memory card <input type="checkbox"/> CD-RW <input type="checkbox"/> Floppy 3.5" <input type="checkbox"/> DVD+RW <input type="checkbox"/> Memory stick <input type="checkbox"/> miniCD <input type="checkbox"/> Floppy 5.25" <input type="checkbox"/> DVD-RW <input type="checkbox"/> External hard drive <input type="checkbox"/> Zip disk <input type="checkbox"/> Floppy 8" <input type="checkbox"/> miniDVD <input type="checkbox"/> Other <input type="checkbox"/>	Lightly scratched	50	GB

Label markings	Brand	E-Rec ID
Cy's Birthday Bash 2012 - copy for Lovell	TDK	N/A

MS#	Collection Name	Box	Folder	CD-R	CD-ROM	CD-RW	miniCD	Zip disk	Blu-Ray	miniBlu-Ray	Floppy 3.5"	Floppy 5.25"	Floppy 8"	DVD+R	DVD-R	DVD+RW	DVD-RW	miniDVD	USB flash drive	Memory card	Memory Stick	ext. hard drive	Other	Notes	ERec ID	JumpIn UID (JIMS#)
1000	Cy (Mascot) Papers	5	22			2			1										2							28-32

# APPENDIX 3

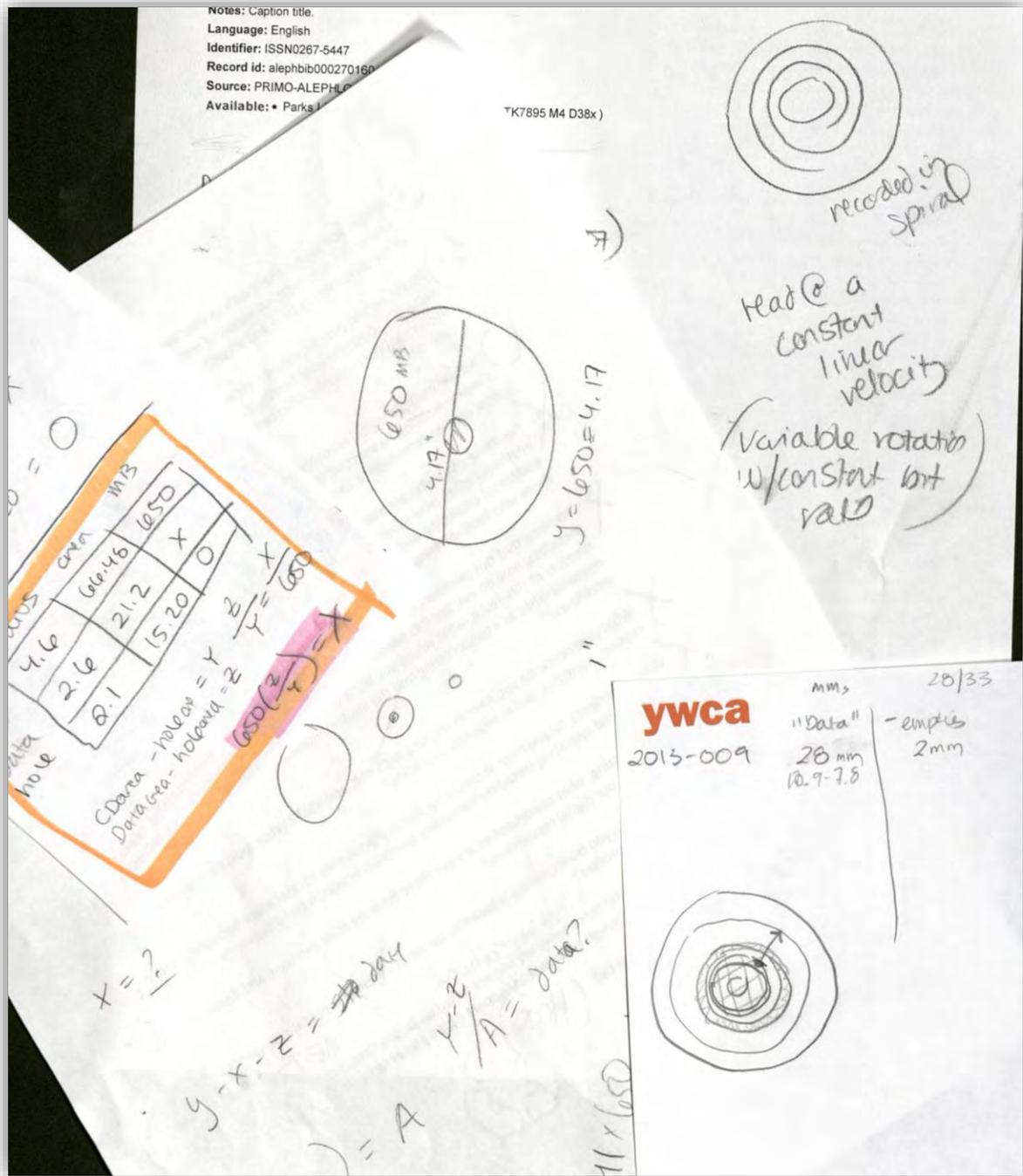


FIGURE 2 - Attempts at figuring out the geometry of the data area on a disk