

Duplicate Checking With Tableau

Brandon Cole and Lloyd Chittenden

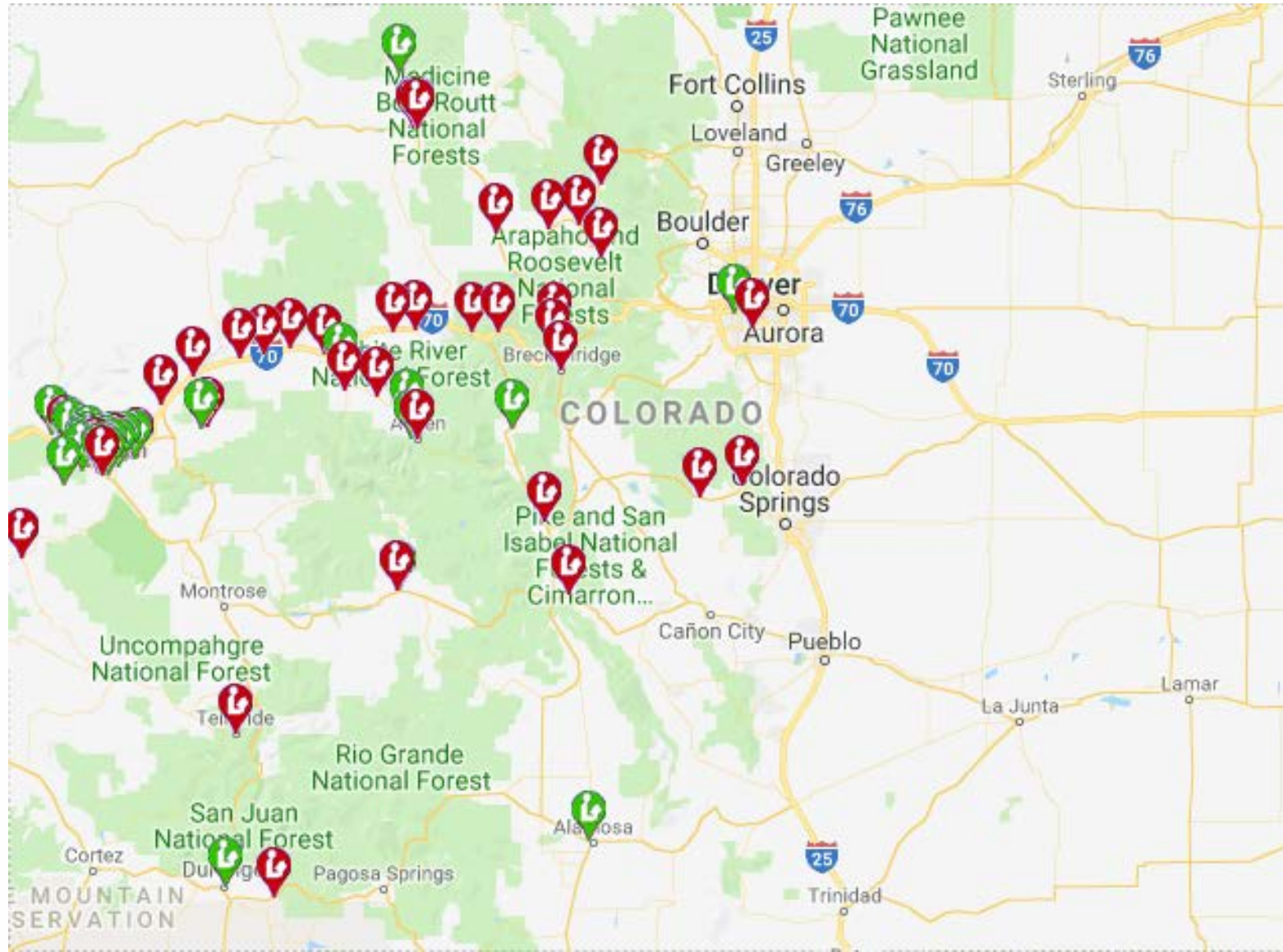


What is Marmot?

- We are a consortium of 29 libraries sharing a single Sierra server
- 19 Public Libraries
- Six Academics
- Four School Districts



We are all over Colorado



What is the problem?

- We have a big problem with duplicates
- We were not happy with Sierra's Headings Reports for finding duplicates
- We realized that SQL allowed us to use Tableau to create our own utility to find duplicates



Why are duplicates a problem?

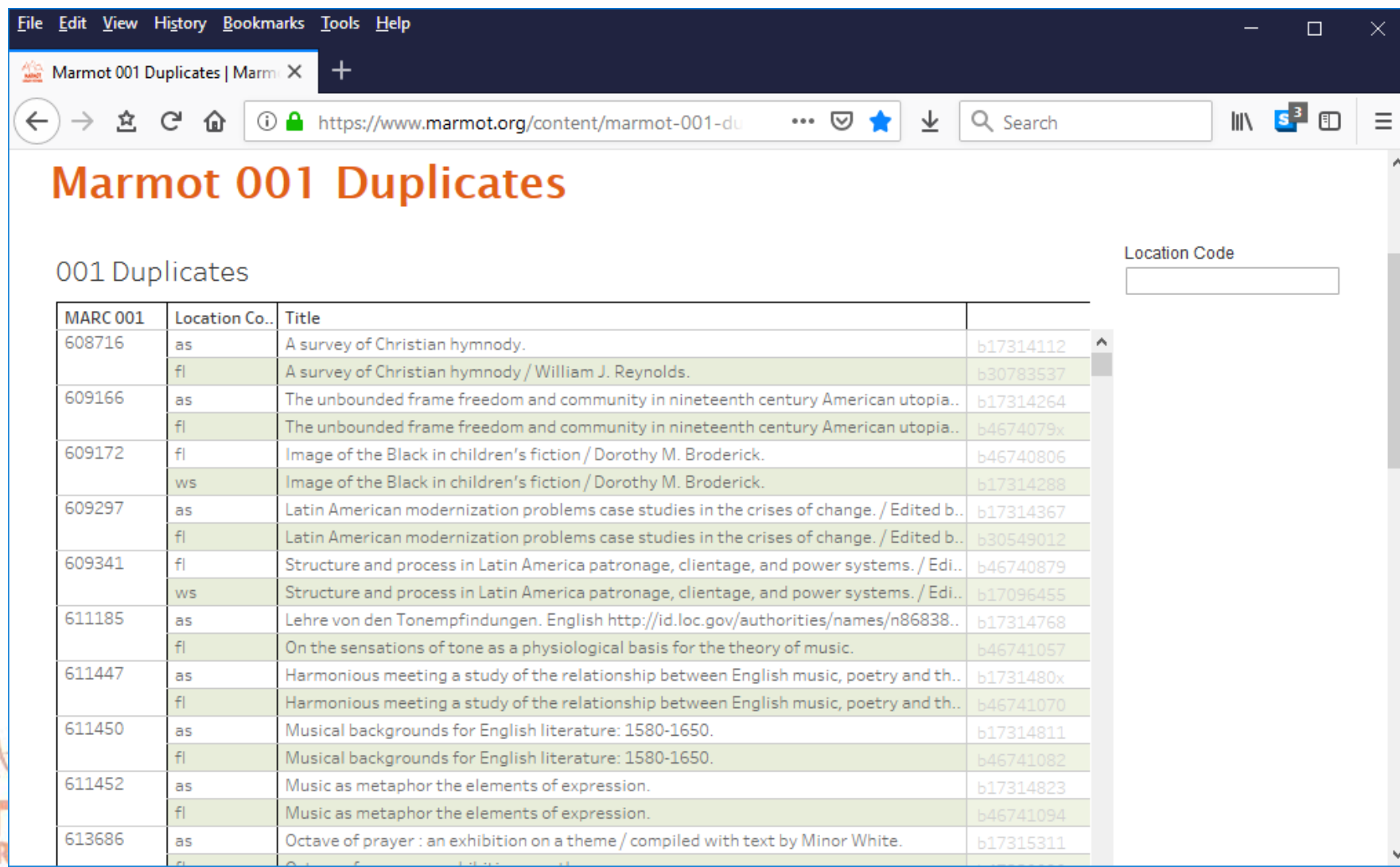
- We have to pay for authority work on each record
- They make holds and requests inefficient



Marmot's solution

- When we started using Tableau and SQL to create statistics reports we realized that it could be used to create a duplicate searching utility
- We created three tools to find duplicates
 - BIB UTIL# (001 and 019)
 - ISN (020, 022, 024, 028)
 - 020a

The duplicate Bib Util # tool

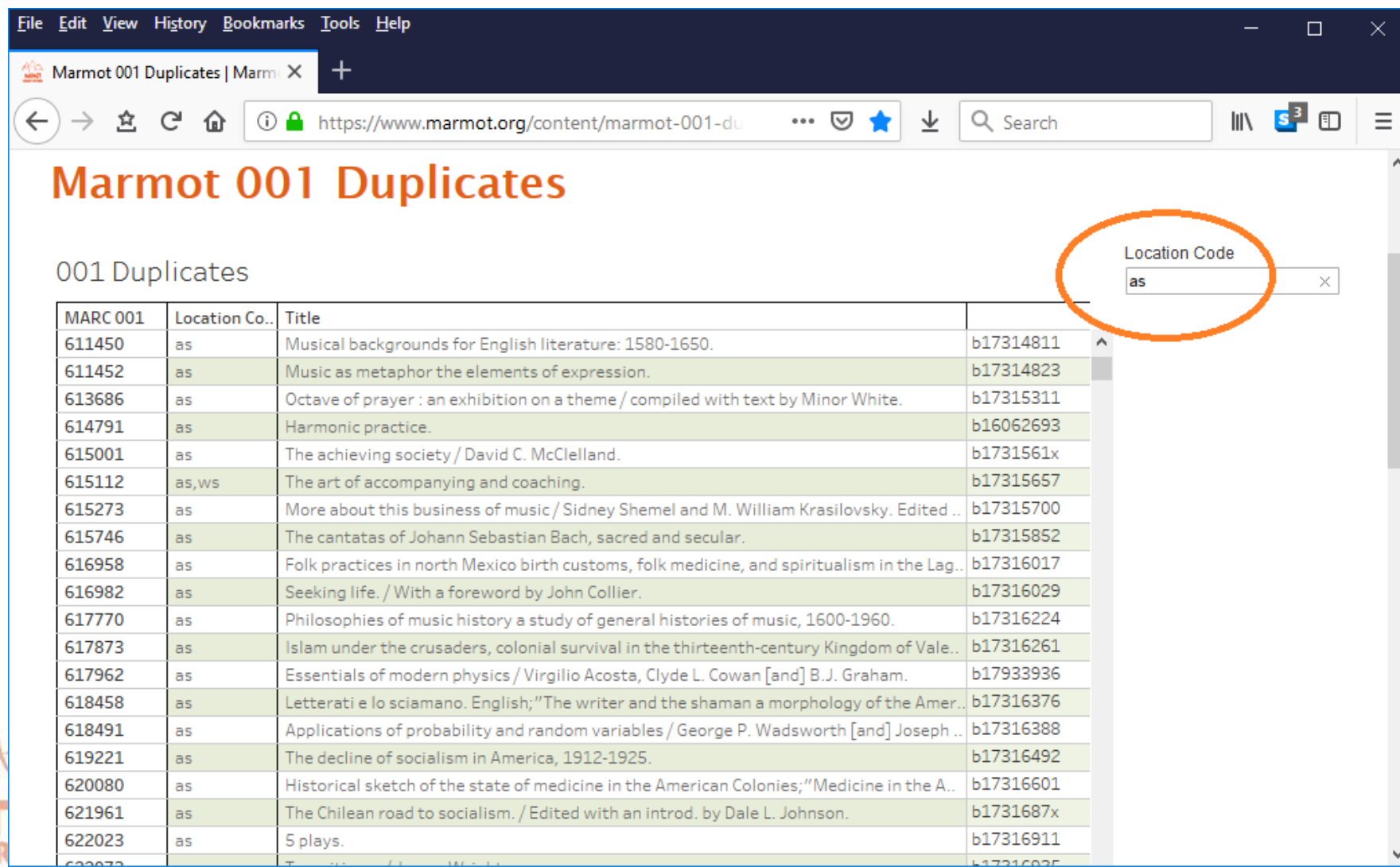


The screenshot shows a web browser window with the URL <https://www.marmot.org/content/marmot-001-du>. The page title is "Marmot 001 Duplicates". Below the title, there is a search bar labeled "001 Duplicates" and a "Location Code" input field. The main content is a table with four columns: "MARC 001", "Location Co..", "Title", and a fourth column containing identification numbers. The table lists 24 rows of duplicate records, each with a unique MARC 001 number and a corresponding location code and title.

MARC 001	Location Co..	Title	
608716	as	A survey of Christian hymnody.	b17314112
	fi	A survey of Christian hymnody / William J. Reynolds.	b30783537
609166	as	The unbounded frame freedom and community in nineteenth century American utopia..	b17314264
	fi	The unbounded frame freedom and community in nineteenth century American utopia..	b4674079x
609172	fi	Image of the Black in children's fiction / Dorothy M. Broderick.	b46740806
	ws	Image of the Black in children's fiction / Dorothy M. Broderick.	b17314288
609297	as	Latin American modernization problems case studies in the crises of change. / Edited b..	b17314367
	fi	Latin American modernization problems case studies in the crises of change. / Edited b..	b30549012
609341	fi	Structure and process in Latin America patronage, clientage, and power systems. / Edi..	b46740879
	ws	Structure and process in Latin America patronage, clientage, and power systems. / Edi..	b17096455
611185	as	Lehre von den Tonempfindungen. English http://id.loc.gov/authorities/names/n86838..	b17314768
	fi	On the sensations of tone as a physiological basis for the theory of music.	b46741057
611447	as	Harmonious meeting a study of the relationship between English music, poetry and th..	b1731480x
	fi	Harmonious meeting a study of the relationship between English music, poetry and th..	b46741070
611450	as	Musical backgrounds for English literature: 1580-1650.	b17314811
	fi	Musical backgrounds for English literature: 1580-1650.	b46741082
611452	as	Music as metaphor the elements of expression.	b17314823
	fi	Music as metaphor the elements of expression.	b46741094
613686	as	Octave of prayer : an exhibition on a theme / compiled with text by Minor White.	b17315311



Each library can limit to their location



The screenshot shows a web browser window with the URL <https://www.marmot.org/content/marmot-001-du>. The page title is "Marmot 001 Duplicates". Below the title, there is a search filter for "Location Code" with the value "as" entered. The filter is highlighted with an orange circle. Below the filter is a table of duplicates.

MARC 001	Location Co..	Title	
611450	as	Musical backgrounds for English literature: 1580-1650.	b17314811
611452	as	Music as metaphor the elements of expression.	b17314823
613686	as	Octave of prayer : an exhibition on a theme / compiled with text by Minor White.	b17315311
614791	as	Harmonic practice.	b16062693
615001	as	The achieving society / David C. McClelland.	b1731561x
615112	as,ws	The art of accompanying and coaching.	b17315657
615273	as	More about this business of music / Sidney Shemel and M. William Krasilovsky. Edited ..	b17315700
615746	as	The cantatas of Johann Sebastian Bach, sacred and secular.	b17315852
616958	as	Folk practices in north Mexico birth customs, folk medicine, and spiritualism in the Lag..	b17316017
616982	as	Seeking life. / With a foreword by John Collier.	b17316029
617770	as	Philosophies of music history a study of general histories of music, 1600-1960.	b17316224
617873	as	Islam under the crusaders, colonial survival in the thirteenth-century Kingdom of Vale..	b17316261
617962	as	Essentials of modern physics / Virgilio Acosta, Clyde L. Cowan [and] B.J. Graham.	b17933936
618458	as	Letterati e lo sciamano. English;"The writer and the shaman a morphology of the Amer..	b17316376
618491	as	Applications of probability and random variables / George P. Wadsworth [and] Joseph ..	b17316388
619221	as	The decline of socialism in America, 1912-1925.	b17316492
620080	as	Historical sketch of the state of medicine in the American Colonies;"Medicine in the A..	b17316601
621961	as	The Chilean road to socialism. / Edited with an introd. by Dale L. Johnson.	b1731687x
622023	as	5 plays.	b17316911
622073	as	...	b17316925

Now we only see dups that include that one library

Marmot 001 Duplicates

001 Duplicates

Location Code
as

MARC 001	Location Co..	Title	
611450	as	Musical backgrounds for English literature: 1580-1650.	b17314811
611452	as	Music as metaphor the elements of expression.	b17314823
613686	as	Octave of prayer : an exhibition on a theme / compiled with text by Minor White.	b17315311
614791	as	Harmonic practice.	b16062693
615001	as	The achieving society / David C. McClelland.	b1731561x
615112	as,ws	The art of accompanying and coaching.	b17315657
615273	as	More about this business of music / Sidney Shemel and M. William Krasilovsky. Edited ..	b17315700
615746	as	The cantatas of Johann Sebastian Bach, sacred and secular.	b17315852
616958	as	Folk practices in north Mexico birth customs, folk medicine, and spiritualism in the Lag..	b17316017
616982	as	Seeking life. / With a foreword by John Collier.	b17316029
617770	as	Philosophies of music history a study of general histories of music, 1600-1960.	b17316224
617873	as	Islam under the crusaders, colonial survival in the thirteenth-century Kingdom of Vale..	b17316261
617962	as	Essentials of modern physics / Virgilio Acosta, Clyde L. Cowan [and] B.J. Graham.	b17933936
618458	as	Letterati e lo sciamano. English;"The writer and the shaman a morphology of the Amer..	b17316376
618491	as	Applications of probability and random variables / George P. Wadsworth [and] Joseph ..	b17316388
619221	as	The decline of socialism in America, 1912-1925.	b17316492
620080	as	Historical sketch of the state of medicine in the American Colonies;"Medicine in the A..	b17316601
621961	as	The Chilean road to socialism. / Edited with an introd. by Dale L. Johnson.	b1731687x
622023	as	5 plays.	b17316911
622073	as	...	b17316935

Copy the 001

Marmot 001 Duplicates

001 Duplicates

Location Code: as

MARC 001	Location Co..	Title	
611450	as	Musical backgrounds for English literature: 1580-1650.	b17314811
611452		pr the elements of expression.	b17314823
613686		: an exhibition on a theme / compiled with text by Minor White.	b17315311
614791		e.	b16062693
615001		McClelland.	b1731561x
615112	as,ws	ching.	b17315657
615273	as	c / Sidney Shemel and M. William Krasilovsky. Edited ..	b17315700
615746	as	n Bach, sacred and secular.	b17315852
616958	as	th customs, folk medicine, and spiritualism in the Lag..	b17316017
616982	as	y John Collier.	b17316029
617770	as	tudy of general histories of music, 1600-1960.	b17316224
617873	as	al survival in the thirteenth-century Kingdom of Vale..	b17316261
617962	as	Essentials of modern physics / Virgilio Acosta, Clyde L. Cowan [and] B.J. Graham.	b17933936
618458	as	Letterati e lo sciamano. English;"The writer and the shaman a morphology of the Amer..	b17316376
618491	as	Applications of probability and random variables / George P. Wadsworth [and] Joseph ..	b17316388
619221	as	The decline of socialism in America, 1912-1925.	b17316492



Paste into Sierra and fix the duplicate

The screenshot shows the Sierra library catalog interface. The search criteria are set to 'BIB UTIL NO' with the value '611450'. The search results table is as follows:

#	Description	# of Entries
1	611450	2
2	→ Musical backgrounds for English literature: 1580-1650. · ASU Nielsen Library C384 · Finney, Gretch...	
3	→ Musical backgrounds for English literature: 1580-1650. · Fort Lewis College C504 · Finney, Gretche...	
4	+ 61145013	1
5	+ 61145093	1

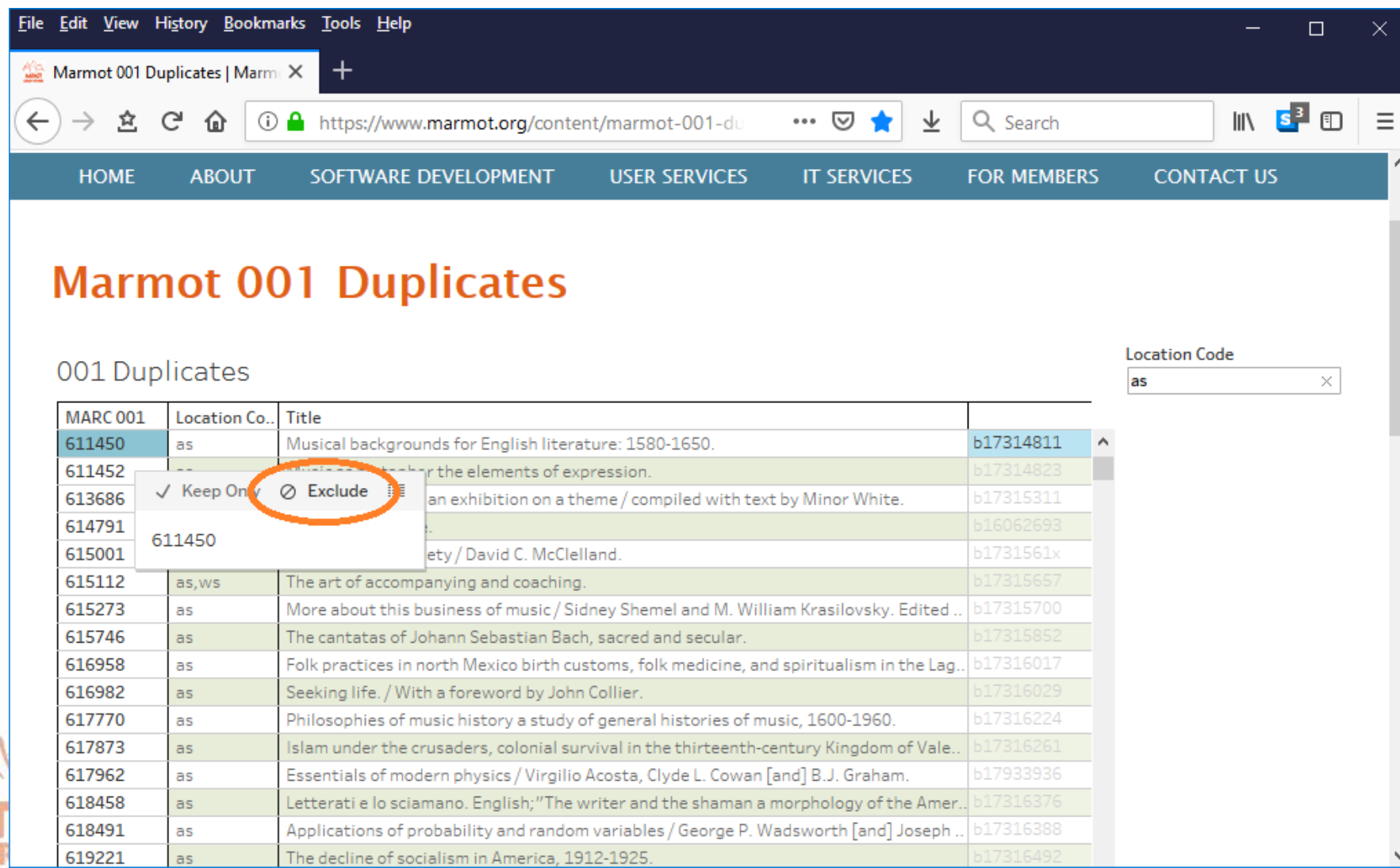
On the left side, there is a filter panel with the following categories and counts:

- Format: BOOK/SERIAL (3), AUDIO BOOK (1)
- Language: English (4)
- Locations: ASU Children's C384 (1), ASU Nielsen Library C384 (2), Fort Lewis College C504 (1), Pitkin County Library C576 (1), SSCL Bud Werner Library C722 (1), Steamboat Springs Libs (1), multi (2)

At the bottom right, a summary states: 3 BIB UTIL NOS, 4 ENTRIES.



Exclude this pair

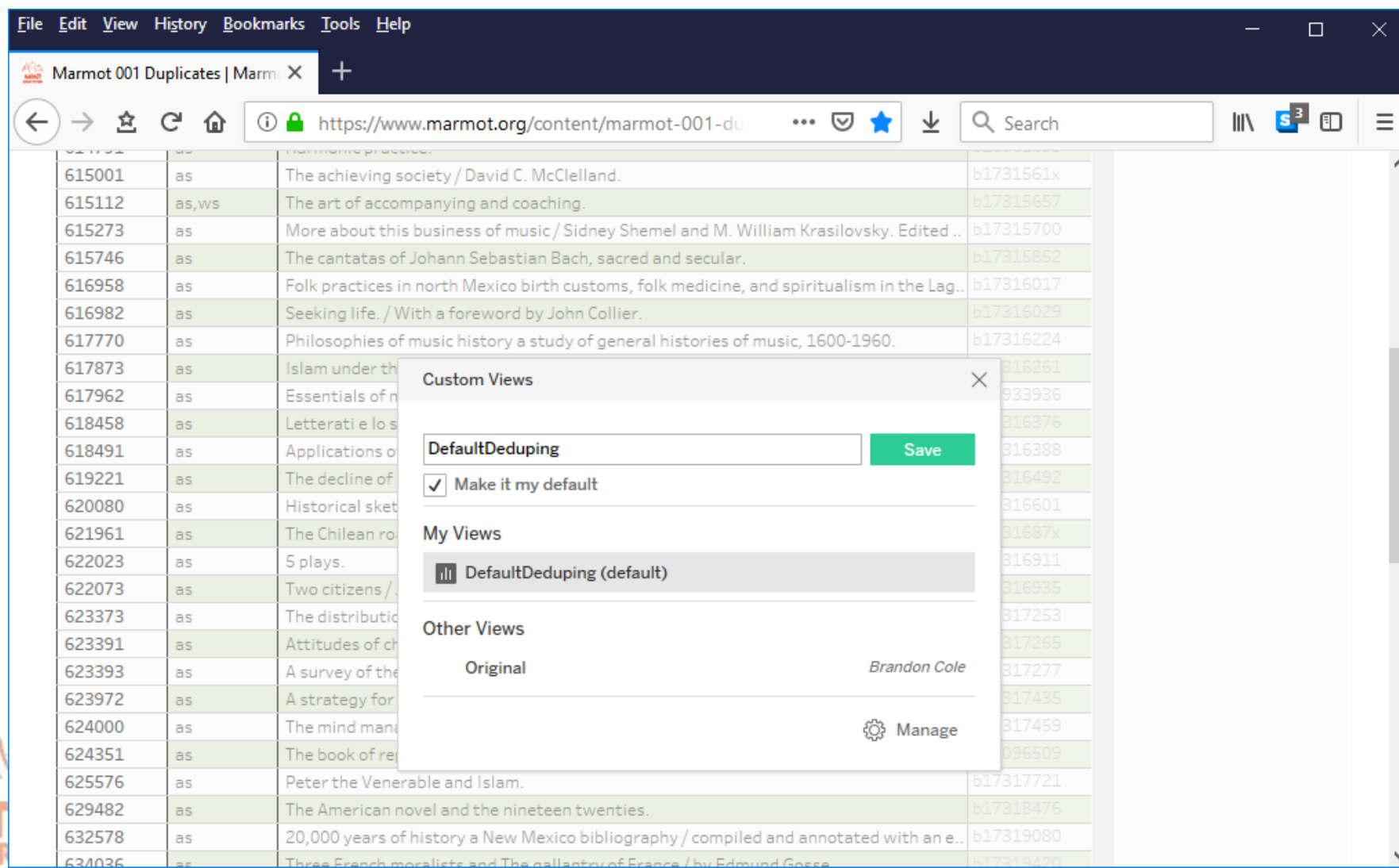


The screenshot shows a web browser window with the URL <https://www.marmot.org/content/marmot-001-du>. The page title is "Marmot 001 Duplicates". Below the title, there is a section labeled "001 Duplicates" and a "Location Code" input field containing "as". A table of duplicates is displayed with columns for MARC 001, Location Co., Title, and a unique identifier. A context menu is open over the row with MARC 001 613686, and the "Exclude" option is circled in orange.

MARC 001	Location Co..	Title	
611450	as	Musical backgrounds for English literature: 1580-1650.	b17314811
611452		... the elements of expression.	b17314823
613686	✓ Keep Only	... an exhibition on a theme / compiled with text by Minor White.	b17315311
614791		...	b16062693
615001	611450	... ety / David C. McClelland.	b1731561x
615112	as,ws	The art of accompanying and coaching.	b17315657
615273	as	More about this business of music / Sidney Shemel and M. William Krasilovsky. Edited ..	b17315700
615746	as	The cantatas of Johann Sebastian Bach, sacred and secular.	b17315852
616958	as	Folk practices in north Mexico birth customs, folk medicine, and spiritualism in the Lag..	b17316017
616982	as	Seeking life. / With a foreword by John Collier.	b17316029
617770	as	Philosophies of music history a study of general histories of music, 1600-1960.	b17316224
617873	as	Islam under the crusaders, colonial survival in the thirteenth-century Kingdom of Vale..	b17316261
617962	as	Essentials of modern physics / Virgilio Acosta, Clyde L. Cowan [and] B.J. Graham.	b17933936
618458	as	Letterati e lo sciamano. English;"The writer and the shaman a morphology of the Amer..	b17316376
618491	as	Applications of probability and random variables / George P. Wadsworth [and] Joseph ..	b17316388
619221	as	The decline of socialism in America, 1912-1925.	b17316492



Save this view to save your excluded dup



The screenshot shows a web browser window with the URL <https://www.marmot.org/content/marmot-001-du>. The page displays a table of book records. A 'Custom Views' dialog box is open, allowing the user to save a new view named 'DefaultDeduping'. The dialog box includes a 'Save' button, a 'Make it my default' checkbox, and a 'Manage' button. The table contains the following data:

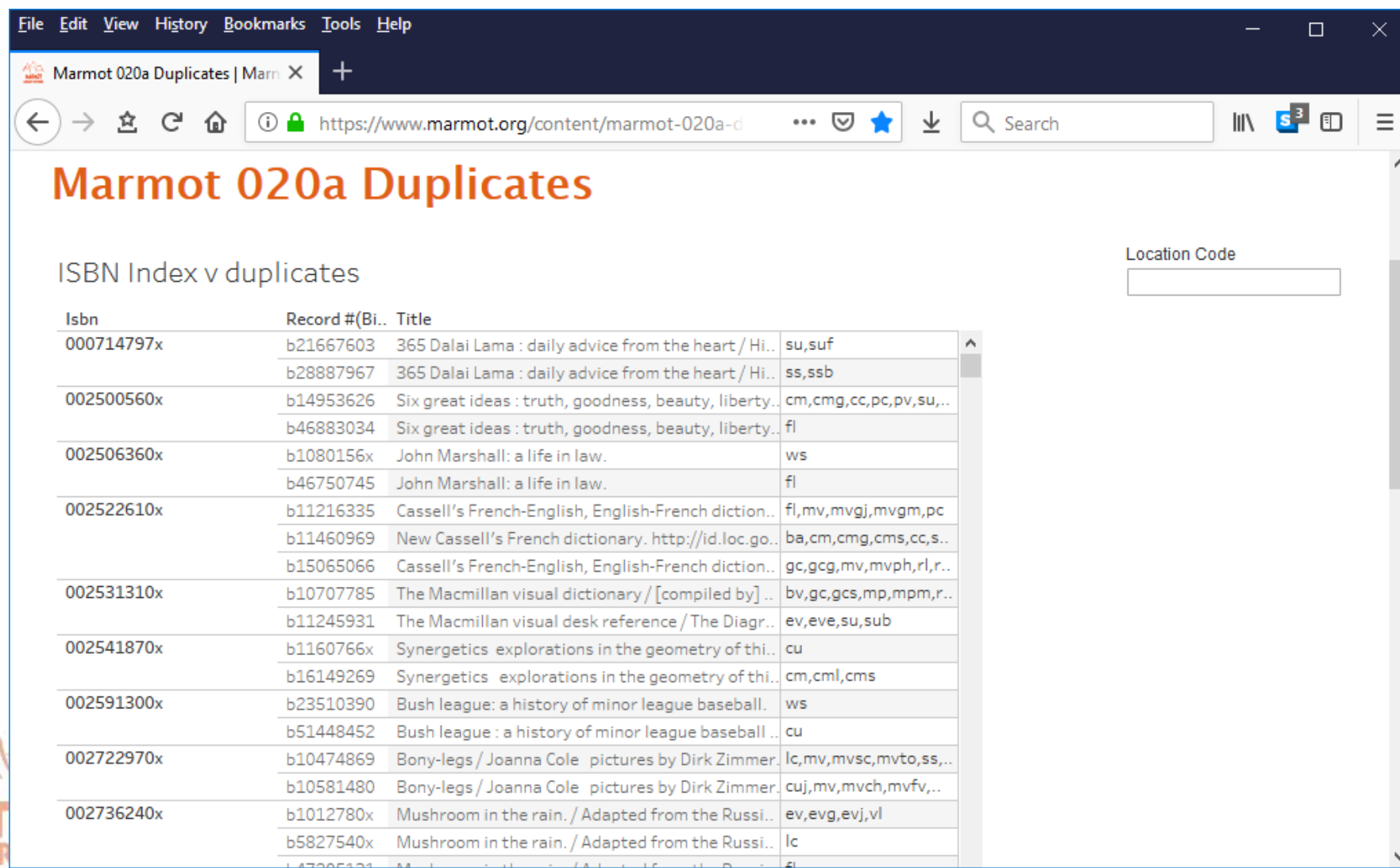
ID	Author	Title	Barcode
615001	as	The achieving society / David C. McClelland.	b1731561x
615112	as,ws	The art of accompanying and coaching.	b17315657
615273	as	More about this business of music / Sidney Shemel and M. William Krasilovsky. Edited ..	b17315700
615746	as	The cantatas of Johann Sebastian Bach, sacred and secular.	b17315852
616958	as	Folk practices in north Mexico birth customs, folk medicine, and spiritualism in the Lag..	b17316017
616982	as	Seeking life. / With a foreword by John Collier.	b17316029
617770	as	Philosophies of music history a study of general histories of music, 1600-1960.	b17316224
617873	as	Islam under th	b16261
617962	as	Essentials of n	933936
618458	as	Letterati e lo	816376
618491	as	Applications o	816388
619221	as	The decline of	816492
620080	as	Historical sket	816601
621961	as	The Chilean ro	81687x
622023	as	5 plays.	816911
622073	as	Two citizens /	816935
623373	as	The distributio	817253
623391	as	Attitudes of ch	817265
623393	as	A survey of the	817277
623972	as	A strategy for	817435
624000	as	The mind man	817459
624351	as	The book of re	096509
625576	as	Peter the Venerable and Islam.	b17317721
629482	as	The American novel and the nineteen twenties.	b17318476
632578	as	20,000 years of history a New Mexico bibliography / compiled and annotated with an e..	b17319080
634036	as	Three French moralists and The gallantry of France / by Edmund Gosse	b17319420

One Tableau login for all dedupers

- Everyone uses the same login, so they all see the same view
- A duplicate pair that one library excludes disappears for everyone
- Each user can still limit to their location within the view
- Tableau allows multiple users at the same time



We also have one that looks at ISBNs



The screenshot shows a web browser window with the address bar displaying <https://www.marmot.org/content/marmot-020a-d>. The page title is "Marmot 020a Duplicates". Below the title, there is a section titled "ISBN Index v duplicates" and a "Location Code" search box.

Isbn	Record #(Bi..	Title	
000714797x	b21667603	365 Dalai Lama : daily advice from the heart / Hi..	su,suf
	b28887967	365 Dalai Lama : daily advice from the heart / Hi..	ss,ssb
002500560x	b14953626	Six great ideas : truth, goodness, beauty, liberty..	cm,cmg,cc,pc,pv,su,..
	b46883034	Six great ideas : truth, goodness, beauty, liberty..	fl
002506360x	b1080156x	John Marshall: a life in law.	ws
	b46750745	John Marshall: a life in law.	fl
002522610x	b11216335	Cassell's French-English, English-French diction..	fl,mv,mvgj,mvgm,pc
	b11460969	New Cassell's French dictionary. http://id.loc.go..	ba,cm,cmg,cms,cc,s..
	b15065066	Cassell's French-English, English-French diction..	gc,gcg,mv,mvph,rl,r..
002531310x	b10707785	The Macmillan visual dictionary / [compiled by] ..	bv,gc,gcs,mp,mpm,r..
	b11245931	The Macmillan visual desk reference / The Diagr..	ev,eve,su,sub
002541870x	b1160766x	Synergetics explorations in the geometry of thi..	cu
	b16149269	Synergetics explorations in the geometry of thi..	cm,cml,cms
002591300x	b23510390	Bush league: a history of minor league baseball.	ws
	b51448452	Bush league : a history of minor league baseball ..	cu
002722970x	b10474869	Bony-legs / Joanna Cole pictures by Dirk Zimmer.	lc,mv,mvsc,mvto,ss,..
	b10581480	Bony-legs / Joanna Cole pictures by Dirk Zimmer.	cuj,mv,mvch,mvfv,..
002736240x	b1012780x	Mushroom in the rain. / Adapted from the Russi..	ev,evg,evj,vl
	b5827540x	Mushroom in the rain. / Adapted from the Russi..	lc



How we built the tool

1. Standard SQL Queries pull the initial list of bib record numbers
2. We upload the bib numbers into a Sierra review file so we can export locations
 - a. SQL queries will pull multiple locations in a way that will appear to be duplicate records, so we have to pull the locations out of Create Lists
 - b. This step may not be needed for single site institutions
3. Copy the locations export into Google sheet
4. Configure and Display in Tableau
5. Display based on a Google sheet also improves performance



001 duplicates SQL query

```
select dups.record_type_code || dups.record_num  
from sierra_view.varfield_view dups  
join (select count(*), field_content, record_type_code  
from sierra_view.varfield_view  
where marc_tag = '001' and record_type_code = 'b'  
group by record_type_code, field_content  
having count(*) > 1  
) seed on seed.field_content = dups.field_content and seed.record_type_code = dups.record_type_code  
order by dups.field_content
```



020 duplicates query

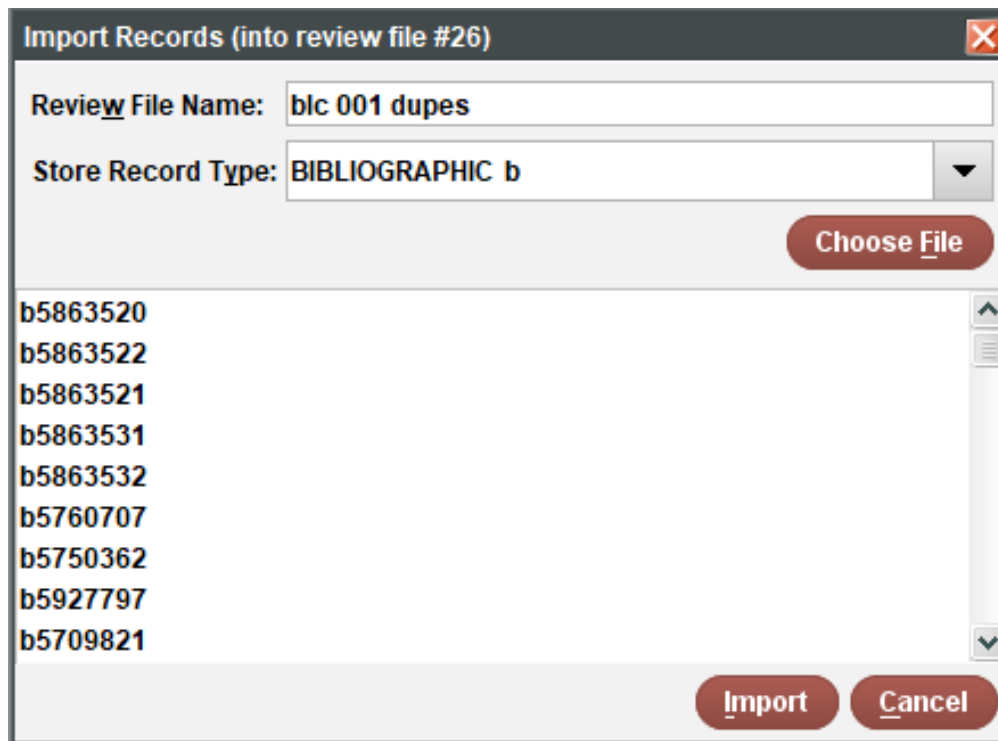
```
SELECT
  id2reckey (p.record_id) || 'a' AS rid,
  p.index_entry
FROM
  sierra_view.phrase_entry AS p
  JOIN sierra_view.bib_record AS b ON p.record_id = b.record_id
WHERE
  --presently limited to e-books
  p.index_tag = 'i'
  AND p.index_entry IN (
    SELECT
      p.index_entry
    FROM
      sierra_view.phrase_entry AS p
    WHERE
      p.index_tag = 'i'
    GROUP BY
      p.index_entry
    HAVING
      count(p.id) > 1)
ORDER BY
  index_entry,
  rid
```

Just need to change these two
index tags to get a different Sierra
index



Import into Sierra

- Since we have bib record numbers we can use “Import Records” in Create Lists
- If your files of duplicates are very large you may need to split them up



The screenshot shows a dialog box titled "Import Records (into review file #26)". It contains the following fields and controls:

- Review File Name:** A text input field containing "blc 001 dupes".
- Store Record Type:** A dropdown menu currently set to "BIBLIOGRAPHIC b".
- Choose File:** A red button to the right of the Store Record Type dropdown.
- Record List:** A list box containing the following bib record numbers:
 - b5863520
 - b5863522
 - b5863521
 - b5863531
 - b5863532
 - b5760707
 - b5750362
 - b5927797
 - b5709821
- Import:** A red button at the bottom right.
- Cancel:** A red button at the bottom right, next to the Import button.

Export from Sierra

- Fields to Export for the Google sheet
 - Record Number
 - 001 or 020 tag
 - Bib Locations Field
 - Bib Location Field
 - Title
- Use something uncommon for the field delimiter like a ~

bic 001 dupes

Export BIBLIOGRAPHIC Information

Fields to be exported

Line	Type	Field
1	BIBLIOGRAPHIC	RECORD #
2	BIBLIOGRAPHIC	MARC Tag 001
3	BIBLIOGRAPHIC	LOCATIONS
4	BIBLIOGRAPHIC	LOCATION
5	BIBLIOGRAPHIC	TITLE

Append

Insert

Delete

Field delimiter +

Text qualifier "

Repeated field delimiter ;

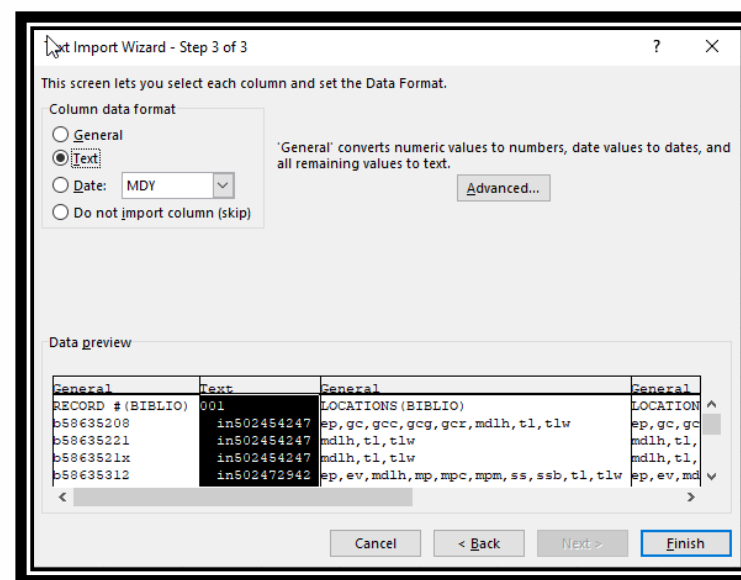
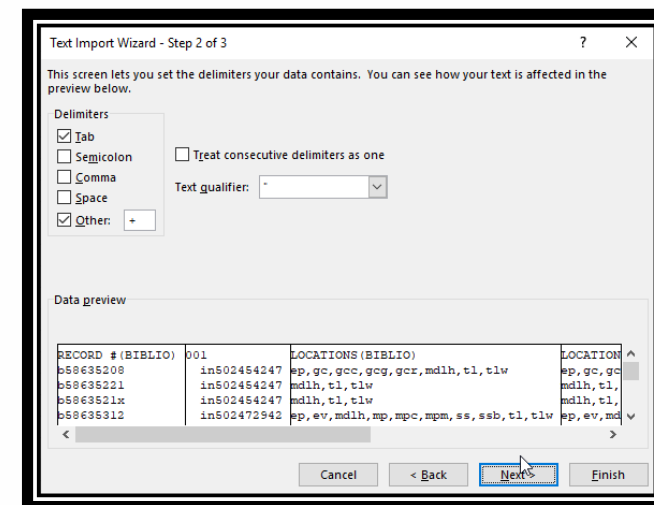
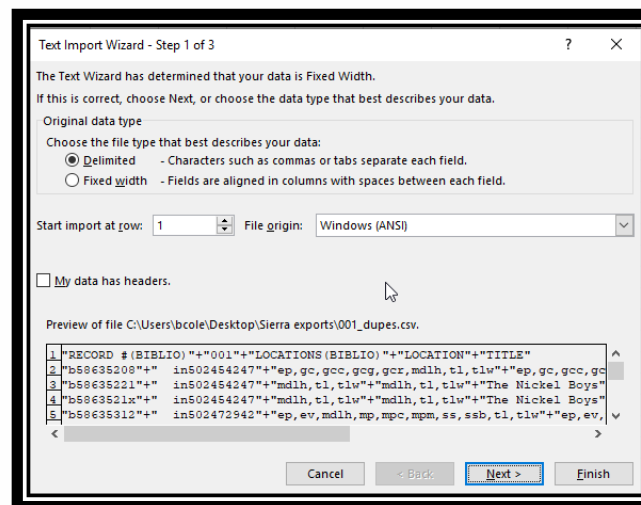
Maximum field length (0-1000) <none>

File: Browse

OK Apply Saved Export Save This Export Close

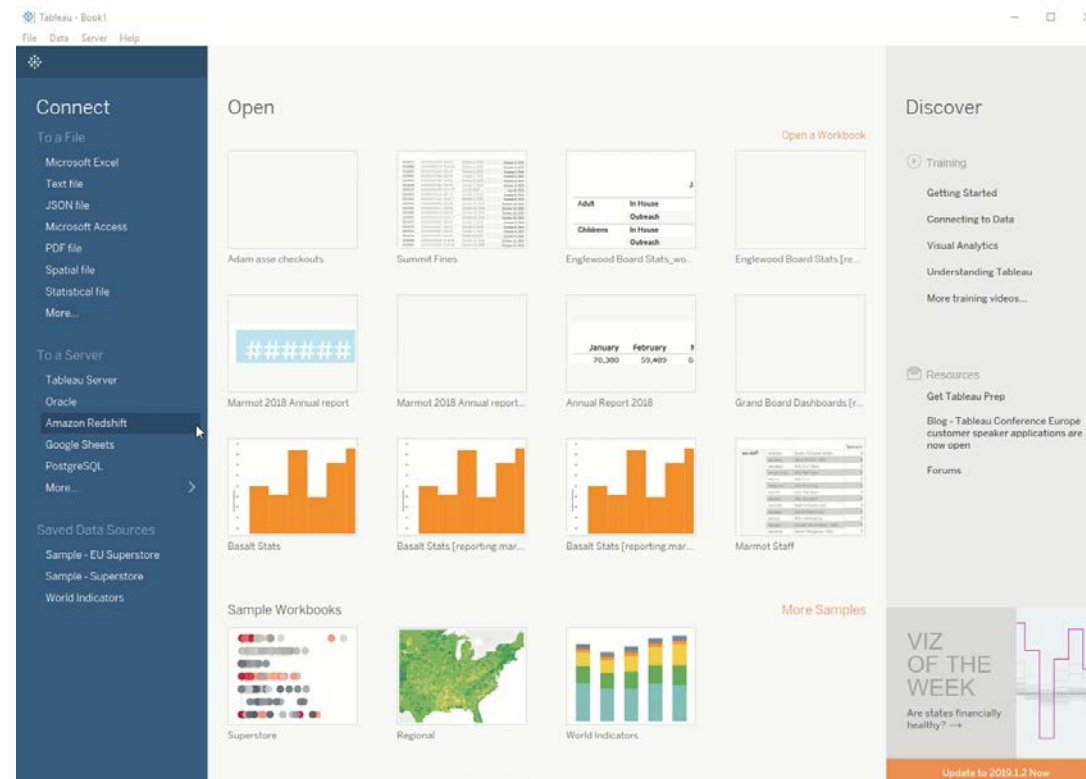
Transfer CSV to Google Sheets

- Best to open in Excel first using Data>From Text
 - Opens a Text import wizard
 - Choose Delimited
 - Choose Other for Delimiter and put in the character you used
 - For the Column data format make sure your 001 or 020 column is formatted as text



Building the Tableau Worksheet

- Connect to the Data Source (Google sheet or Excel)
 - For Google sheets this will open a browser window to choose the google account
 - Next choose the document you want to link to



Building the Tableau Worksheet

- Pull the data into the rows
 - Depending on the amount of data, the worksheet can take a long time to update upon adding each row
 - Add record number as a mark so that last column isn't empty
 - Add desired filters (location)

Tableau - Marmot Duplicates

File Data Worksheet Dashboard Story Analysis Map Format Server Window Help

Standard

Columns: MARC 001, Location Code

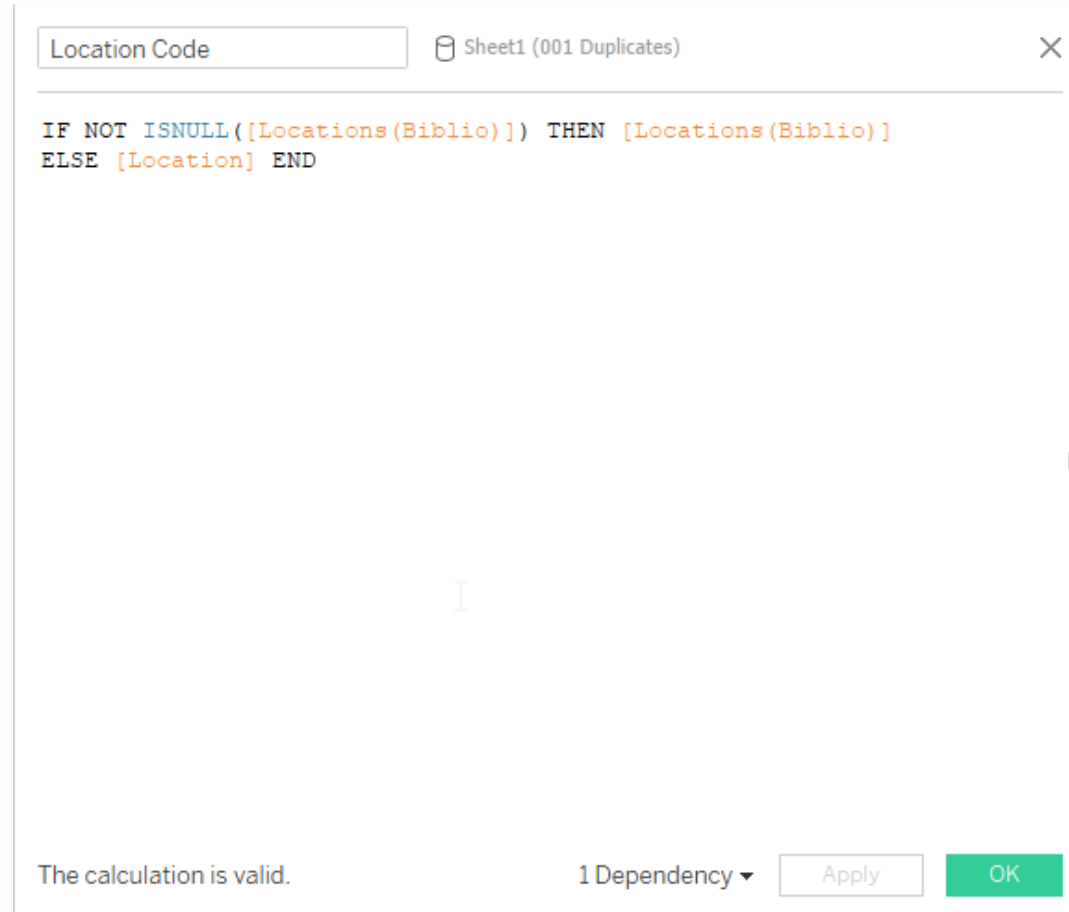
Rows: 001 Duplicates

MARC 001	Location Code	Record # (Bibli)
357354:"148109198"	as,asju,cuj,ev,evb,evg,evj	b10072615
607481	as	b1606222x
608094	fl	b46740648
608356	cm,cmg	b16062255
608716	fl	b4694039x
609166	cc	b30515701
609172	cm,cmj,ws	b16009721
609297	as	b17314288
609341	fl	b17314367
611185	fl	b30549012
611447	ws	b46740879
611452	fl	b17096455
613686	as	b17314768
614178	fl	b46741057
614791	as	b1731480x
615001	fl	b46741070
615273	as	b17314823
615631	fl	b46741094
615746	as	b17315311
616958	fl	b47330090
	ws	b46741550
	as	b16062644
	cc	b16062693
	as	b30686854
	cc	b1731561x
	as	b31038761
	fl	b17315700
	fl	b46741677
	pc	b4694073x
	as	b15707210
	fl	b17315852
	fl	b47117928
	as	b17316017

16042 marks 16012 rows by 1 column

Building the Tableau Worksheet

- A calculated field in Tableau combines the locations and location field if the locations field is empty



Building the Tableau Worksheet

- From here you can either export to a Tableau server or share the workbook locally
 - Tableau server is a separate subscription.
 - Allows password protection web access
 - Can share locally as a workbook to others who also have Tableau Desktop

Live Demo

