Web of Science field	Niner Commons field	OpenRefine data transformations	Manual interventions still needed
Article Title	Title	Split the Article Title Column using ": " (colon-space) as the separator Transform using **value[0].toUppercase()+value.toLowercase().substring(1,value.length())** Rejoin the column using " : " (space-colon-space) as the separator	Review for proper nouns and acronyms
[n/a]	UNCC Constituent Type	n/a	Enter manually
Author Full Name	Author: Type	n/a	set to "personal"
	Author: Controlled Vocabulary Used	n/a	set to "Local File or No Authority File Found" or "LC Name Authority File" if author is listed in LCNAF
	Author: Name	Split Author Full Names column using split multi-valued cells with "; " as the separator. Cluster and edit the column. Use key collision/fingerprint then nearest neighbor/ppm. Re-join the column cells using ";" as the separator. Split the column into several columns using ";" as the separator. Split into 6 columns at most. Uncheck "Remove this column." Delete the 6th column.	Will need to manually review items with many authors.
	Author: UNCC E-mail Address	n/a	Look up in directory
	Author: Assigned Linked ID By Controlled Vocabulary	Reconcile Author Full Name columns using LC VIAF API. Add new column based on Author Full Name column called "LCNAF" using **cell.recon.match.id** Transform new column using **"http://id.loc.gov/authorities/names/" + cells['LCNAF'].value**	In testing, only about 10% of these matched using the auto-match, so there were many to review. It may be easier to look them up manually depending on how many rows are in the file. Consider a numeric facet on the column to focus on the high probability matches.
Reprint Addresses	Author: Department	Add column based on Reprint Addresses column using the following Python/Jython: **import re** **pattern = re.compile(r"((Dept|Sch|School|Department).+?),", re.I)** **list = []** **for i in pattern.findall(value):** ** list.append(i[0])** **return ";".join(list)** Find "Dept" and replace with "Department of" and "Sch" with "School of" Cluster and edit on column (use nearest neighbor and ppm) and create a text facet to clean normalize department names as able. Reference this document for controlled terms.	Look up in directory for any blanks
	Author: ORCID ID	Reconcile Author Full Name column using ORCID API. Add new column based on Author Full Name column called "ORCID" using **cell.recon.match.id** Transform new column using **"https://orcid.org/" + cells['ORCID'].value**	This can also be cumbersome to review. In testing, none would automatch so it requires you to manually review all of them. Consider a numeric facet on the column to focus on the high probability matches.
[n/a]	Author: Role	n/a	Set to "author"
Abstract	Abstract	n/a	Review for typos/formatting issues

Web of Science field	Niner Commons field	OpenRefine data transformations	Manual interventions still needed
Publication Type	Material Type	Create a text facet on column; edit types in facet (J=Journal article; C=Conference proceedings)	
[n/a]	Format/MIME Type	n/a	Set to "application/pdf"
[n/a]	Digital Origin	n/a	Set to "born digital"
[n/a]	Host Institution	n/a	Set to "J. Murrey Atkins Library (University of North Carolina at Charlotte)"
Publication Year	Single Date of Publication (YYYY)	Create a text facet on Publication Year column; select blanks	
		Transform column using **cells['Early Access Date'].value.substring(4)**	
[n/a]	Issuance	n/a	Set to "single unit"
Author Keywords	Subjects	Split the Author Keywords column into several columns using "; " (semicolon-space) as the separator. Split into 4 columns at most. Uncheck "Remove this column."	
		Reconcile first three columns using FAST reconciliation. Delete 4th column.	
		Add new column based on each reconciled column using **cell.recon.match.id**	
Document Type	Genre Terms	Create a text facet and edit document types as needed. They should largely be articles or conference proceedings. Make a copy of the column.	
		Reconcile new column using Getty API: https://www.getty.edu/research/tools/vocabularies/obtain/openrefine.html	
		Add new column based on reconciled column called "Getty URI" using **cell.recon.match.id**	
		Transform new column using **"http://vocab.getty.edu/" + cells['Getty URI'].value**	
Language	Language	n/a	Filter list to identify any not in English and check those against MARC Code list for Languages for authorized term
Source Title	Journal/Book/Host Title	Transform using: **value[0].toUppercase()+value.toLowercase().substring(1,value.length())**	Review for proper nouns and acronyms
DOI	DOI	Facet column by blank; select false	
		Transform using: **"doi:" + cells['DOI'].value**	