[CEAL Comments on PCC Discussion Paper on Authority Records for Undifferentiated Personal Names](http://www.surveymonkey.com/MySurvey_SettingsTitle.aspx?sm=u0R4PRTQpoGVu21ZCuEOr2kd%2fhn1UbgqzTG1WqkMWm0%3d&TB_iframe=true&height=200&width=480)

**Background information**

On March 29, 2012, PCC posted on its listserv the PCC discussion paper "The future of undifferentiated personal name authority records and other implications for PCC authority work." The paper is posted on PCC website: <http://www.loc.gov/aba/pcc/Undiff%20Personal%20NARs%20Discussion%20Paper%20March%202012.doc>

The CEAL Committee on Technical Processing (CTP) and the CJK NACO Project Coordinator (Sarah Elman) agreed that it was necessary to conduct a survey to gather responses from the East Asian library community. On May 14, 2012, the Chair of CTP (Shi Deng) posted an online survey on the CEAL listserv (Eastlib) and the OCLC-CJK listserv to solicit CEAL members’ feedback. The deadline for response was set for May 31, 2012.

**Survey Summary**

A total of 85 people participated in the survey during the three-week period. Among them, 74 people (87%) supported PCC to pursue the break-up of undifferentiated personal name headings (i.e., splitting up authority records for undifferentiated personal names into multiple records, one for each entity) and 11 people (13%) opposed. Among the 66 people who answered the question “Are you a NACO participant?” 47 people (71 %) checked “Yes.”

We received many positive and constructive comments from participants who support the break-up of records as well as many reasonable comments/concerns from those who do not support the break-up. Please read Appendix II for all the original comments received. The following paragraphs are the summary of the key comments.

Among those who support the break-up, many expressed that this is long overdue and the change will be beneficial for both users and libraries. Meanwhile, many also stressed the importance of having a reliable mechanism in place to distinguish people with identical names, if the break-up project is pursued. Suggested methods include: a) incorporating and leveraging names in CJK scripts in headings or qualifiers; 2) utilizing RDA data elements; and 3) utilizing unique data that already exist in 670s of undifferentiated name records, etc. Some also pointed out that the discussion paper didn’t address CJK specific issues. They recommended that East Asian community be involved in the process, and that CJK catalogers should actively participate in national-wide review and cleanup projects as needed.

For those who do not support the break-up, some pointed out that having undifferentiated NARs is still a viable option under the current environment and cautioned that we should wait until after RDA implementation or after some issues have been addressed--such as undifferentiated NARs lacking paired 670s or having readily differentiated data; NARs coded differentiated but having undifferentiated names in the records, etc. Some thought the break-up is very impractical for searching because users might need to scrutinize among many identical headings in OPAC to identifying the right person they are looking for. Others were also concerned about the scope of its impact on local libraries’ database maintenance and cleanup efforts due to large size of CJK collections as well as shortage of manpower.

These concerns/comments all make valid points, and we hope that PCC will take them into account when making the final decision. The CEAL cataloging community would like to get involved in the planning and implementation of any related project no matter which option is taken. We appreciate the opportunity for discussing the issues and voicing our feedback.

**Appendix I. CEAL Survey Result Summary** (see attached)

**Appendix II. CEAL Survey Individual Comments** (numbers are corresponded to individuals if more than one comments made to these four questions listed below)

**Q2. Comments and/or concerns regarding your decision to support or not support (i.e.: pros and/or cons of splitting up authority records for undifferentiated personal names)**

Support:

|  |  |
| --- | --- |
| 9 | That'd be so beneficial to users and libraries. |
| 10 | I think it's a great idea! Differentiating names as separate identifiable 'entities' rather than spending time on correct creation of a form would be a big time savings. Since 'name form' seems to be most useful in the less-used Browse or Facet searches, rather than the more predominant Keyword searches, I don't think it will have a very major impact on discovery. |
| 18 | Although it involves much work it will increase the identifying and foundabillity of resources. Also increase collocating functions of catalog. |
| 19 | I support the splitting up of authority records for undifferentiated personal names because most of the same personal names are clearly from different historical periods of Chinese history after reading the text of the work. |
| 21 | It's way past the time we should have done this. Technologically it's not a big deal. Also, anyone who has to work with Chinese, German, or Swedish personal names will understand the issue here. |
| 23 | undifferentiated personal names are not helpful. |
| 26 | Support the decision basically, but it will be very time consuming if we have to look up each record for a same name to see which one is right author we are looking for. How about showing either subject or genre of author's work on search result screen or something in that line.... |
| 27 | Undifferentiated personal names result in inaccurate information and confusion. |
| 28 | It never made any sense to me to create “undifferentiated personal name” authority records. |
| 29 | good for users and easy to identify |
| 32 | I strongly support splitting up authority records for undifferentiated personal names. I would not worry too much about how automated systems function in handling the to-be-splitted underfferentiated NARs. My understanding is that the systems (including Connexion Client) currently have not linked the undifferentiated NARs to the bibliographic headings. The RDA NARs, unlike AACR2 NARs, will fulfill the FRAD user tasks with the entire record rather than simply the 1XX field. The approch, if approved, will make much more Chinese NARs differentiated. |
| 35 | too many Chinese authors have same names. |
| 36 | In general, I would like to support the PCC's decision, but in fact, I don't quite know enough to about the pros and cons of this issue. |
| 39 | Is there also any consideration for converting Wade-Giles data fields into the Pinyin scheme in LC/NACO file? Some authority records cover only Wade-Giles data in reference source fields. The maintenance of pinyin data for authority records is critically important to cover Chinese bib. records maintanence. |
| 42 | Perhaps this practice might be easier for catalogers to follow, although it might end up increasing the number of authority records. |
| 47 | i am not a cataloger but it seems to me that if titles written were linked to different names, it would be possible to link the appropriate record to a new book. One would not link a novel or study of history to someone who has written many articles in chemistry. |
| 49 | Different names lumped into one authority record confuse the users. |
| 59 | I support the breaking up of undifferentiated personal names authority records under the condition that a reliable mechanism will be in place to distinguish people with identical names. |
| 61 | Because neither the Wade-Giles nor the Pinyin romanization can replace the function of the original Chinese script. The romanized form word-phonetic transliteration presents no morphological meaning to the Chinese readers. As result, even Chinese-speaking professionals, who are familiar with both subject materials and the above romanization systems, may face difficulties in identifying the correct names. Therefore, it is very important to use the Chinese script for each personal names. |
| 65 | This is difference in the Chinese names especially written in Chinese characters, but the differences are not there when they are being romanized. It would be better that each names has its unique record. However, it may be a bit difficult to distinguish [English or other language] names without dates, unless their occupations, their titles or other characteristics are used. |
| 68 | Whether this benefits or not is depending on how to break-up the record. I would strongly recommend on using vernacular language forms of name to differentiate names which are mixed becaused of romanization. |
| 69 | Undifferentiated personal names are not helpful. we shouldn't creat undiferentiated personal name headings at first place. |
| 70 | Due to machine-derived non-Latin script reference project, many unique personal name headings have different forms of non-Latin names populated into these records, these records may need to be reviewed and broken up as well. Please see examples: LCCN: nr 99032373 Wang, Ke'an (2 different names) LCCN: n 80020091 Hong, Jiaoyi (2 different names) LCCN: n 85009864 Li, Meizhi (4 different names, including one that should be Li, Weizhi) LCCN: no2007124154 Chen, Weiqun (3 different names) |
| 73 | In the past, CJK data was not available in NAF. But now the information can be used to differenciate more individuals even with automatic conversion process. Also we have additional online tools than in the past to differenciate more names for new records. So at least for new records, we have more aids to efficiently achieve this. Cleaning up the local authority file can be very difficult. But we have been seeing small changes like diacritic use, addition of deceased date, etc. and the headings are splitted into so several entries for many authors. It may not be a bad idea to have this nation-wide project which involves not only librarians but also utility managing organizations and library system vendors and do a bit clean-up of local files. |
| 74 | I support this to a certain extend. It is not going to work with just the 3xx fields and fixed field code. Most of the ILS only display the 1xx and 4xx fields of the authority record in the opac. Users very seldom think whether authors are the same or not. To them, the same name is the same name. The splitting may work to authority control aspect but not indexing. It is still 1 name, multiple authors. |
| 82 | I would support to break up undifferentiated personal name headings if they could be identified as different names. If unsure, keeping them on the undifferentiated name record would be still an option. |

Not Support:

|  |  |
| --- | --- |
| 4 | I half-support this. I do not like seeing undifferentiated records with one uniform romanized name with several possible original-script equivalents. On the other hand, years of undifferentiation will be difficult to undo. It would also take a lot of work to differentiate the current records with adequate biochronological data as well with near-exhaustive bibliographies to support the differentiation of authors, who may not even be prolific enough to have birth/death dates readily ascertainable or more than one work associated with them. It seems downright silly to establish an authority record for an author, for example, who wrote one insignificant paperback in the 1910s. In addition, we should rather allow for bibliographic records to serve as surrogate authorities, much like in cases of uniform titles for journals, unless an author is prolific or enigmatic (i.e. multiple aliases) enough to truly require an authority file. The fear amongst catalogers to not input birth or death dates in bibliographic records, even if these dates readily ascertainable from the resource itself or in tandem with other (i.e. non-DLC) national library catalogs, is a joke. Let the PCC empower others, even if these others make mistakes: PCC is not beyond error, as many of my experiences using OCLC have taught me. |
| 15 | My suggestion would simply delete undifferentiated personal name headings. Undifferentiated personal name headings are useless and waste of time for catalogers and NACO members as we need to check each 670 in such a record. Some NACO members choose (or forget) to check and simply establish new NARs with dates. This fact also shows undifferentiated personal name headings are meaningless. |
| 16 | There are too many problems with the existing undifferentiated NARs. Many of them lack additional 670 field pairs, when they cover many different persons. In addition, those NARs coded differentiated have actually been used in corresponding bibliographic records in an undifferentiated way. Furthermore, some undifferentiated NARs are of suspect, carrying readily differentiatable data indicated in existing field 670s. Those problems will need to be resolved before proceeding with this issue any further. |
| 25 | We have a large collection of CJK. I don't think our library system is able to deal with the change. |
| 40 | This would not work the best for CJK languages where multiple scripts or characters are represented in the same romanized form. |
| 45 | increased difficulty to locate the matching heading and confusion with more hits of same name forms. I simply think with all these additional work and headings, it won't change the fact that it won't make much difference in users' experience in searching for names. |
| 58 | I have been trying to imagine how one would make use of a name authority file in which identical names, or names that normalize identically, are represented by individual differentiated NARs, as opposed to the undifferentiated NARs that we have today. What would a user or cataloger encounter when searching in the name authority file? Even though RDA allows considerable latitude in assigning qualifiers in $c subfields, I am not confident that the information found in the 670 fields of the NAR would be helpful in distinguishing any of the individuals listed there (i.e. Resident of U.S.A.? Summer Intern in the Research Dept., IMF? Ph.D. in International Relations?). At any rate, I expect that under RDA most of the 21 persons included in the record for the heading Li Li would still end up being represented by headings that just read Li, Li. Then I imagine that a search for Li, Li in the authority file would result in one of two displays: 1) a list of many identical names, or 2) a single collective entry representing all of the headings reading Li, Li (with a subsequent click leading to a list of some sort). Even if Chinese forms of name are provided, in either case the cataloger or user could be obliged to call up and scrutinize a number of different NARs to locate the person he/she seeks. If that is true, I should think that it would be much more convenient for the cataloger or user to be able to reach all of the information about people represented by the heading Li, Li in one location, on a single, undifferentiated NAR. Think of how much time would be saved over the course of a year. When identifying information can be added to distinguish one of the names, a new unique NAR could be created. The proposal notes that undifferentiated NARs interfere "with the human and machine uses of authority data". I believe that the library world should urge those who wish to make use of the data found in the undifferentiated NARs to modify their programs to make use of the data as it is now structured. 1) Perhaps a program could be devised to deconstruct undifferentiated NARs in such a manner that the VIAF could make use of them -- something along the lines of what is proposed in Appendix C. 2) I do not understand the point that Chinese names could be distinguished by "the original script form of heading in an 880 field. These MARC fields, however, can only be used in authority records for unique entities." Unidfferentiated NARs already include nonroman data in 400 and 670 fields. 3) Intersting idea, but utterly unrealistic. Synchronizing headings between bib and authority records on any but an ongoing basis is impractical. Nobody has the resources to bring a large database into conformance moving forward. Working to synchronize a tiny percentage of the database, but not the rest, would be a terrible waste of time. [see following section] 4) Either the 37x could be programmed to be applied to undifferentiated NARs, so that humanities scholars could embed their data into portions of undiffentiated NARs as well as separate ones, or, the data they add could be used to create a unique heading. I believe that the Resolution in Part Two, B, that "the ID number of the authority record be used as the differentiating characteristic of last resort" will be highly impractical to implement in large and growing databases. One has only to think of catalogers everywhere searching for and using the 21 headings for Li, Li... As for Part Three, this should be the real paradigm shift: Be clear in distinguishing LIBRARY needs from the various possible uses of library data. Retain the current format because it is useful and economical for processing purposes; then devise or modify programs to make use of or extract the data in the undifferentiated records in creative ways to suit the initiatives listed in Part Three. This would save LIBRARIES a great deal of time and money, now and in the future. The proposal cites "connection of headings in bibliographic descriptions to related authority data" as a justification for creating separate NARs for identical headings. As noted above, I believe this to be an impractical goal, particularly for Chinese headings, and therefore not a valid justification for separating undifferentiated NARs. I don't think it is realistic or desirable to plan to implement this linkage in a large database. I considered only how the linkage might be applied to headings for Chinese names. Chinese names present certain challenges. When romanized headings on authority records were converted to pinyin in the year 2000, there were 8,400 undifferentiated NARs for romanized Chinese personal names -- about 30% of all undifferentiated NARs at that time. Of course, the number of undifferentiated NARs for Chinese names will have grown since then. Greatly complicating the identification of Chinese people has been the tendency in recent years for an increasing number of Chinese personal names to consist of just two syllables, a family name and a given name (i.e. Li Li, Liu Liu). Also, many, many people have names using the same characters. I thought of looking at a couple of examples of NARs for undifferentiated Chinese names and walk through how they would be changed from representing more than one person to separate NARs representing single individuals, and then linked to the bib records on which there are entries for the heading. [CTP members: please review these examples and make corrections and improvements where needed. It is likely that the processes I have imagined are out of date, or fail to include certain efficiencies or shortcuts, or I may have overlooked some necessary steps...] Example 1. The record for Li Li (n81113180) represents 21 people. 11 different character combinations are given in the 400 fields. In this case, the NAR conveniently gives the Chinese characters that correspond with each of the distinct names that are cited in 670 fields. Suppose that 21 distinct NARs are created, one for each of the 21 different people named Li Li. 4 of them will not only have the same romanized form, but the same characters 李力. It is likely that one or more of these individuals are represented by a heading on more than one bib record. The only way to accurately link the NAR for one of these 李力s to the records on which his/her heading appears is for a person to painstakingly search each entry in the database, evaluate it, and manually link a bib record to the appropriate 李力 NAR. There are 3 names on the undifferentiated NAR for people whose name is written 李立 in Chinese, and two each for 李丽 and 李莉. There are two persons named Li Li whose work cited in the NAR is in English; it is possible that they are also represented in Chinese form on bib records. In fact, the heading Li, Li (without date) appears in descriptive access points in 84 records in the LC database alone. If one intends to bring order to Li, Li in the OCLC database, the numbers of equivalent names and bib records to evaluate would be far greater, including many records that are not under authority control. To perform authority work, one now refers to a single undifferentiated NAR for all the people whose name is written 李力. However, if each person will be represented by a separate NAR, a cataloger or user may be obliged to call up and peruse four NARs to identify a name or find the NAR for the person whom he/she seeks (cumbersome, to be sure, but not unique -- consider the 23 NARs for persons named John Smith or the 33 NARs for persons named John Williams that do not have corresponding Chinese characters to at least somewhat distinguish individuals). Example 2. The NAR for the heading Chen, Fang (n82091272) covers 6 persons. 10 Chinese forms are given in 400 fields; the first Chinese 400 is clearly incorrect, and the last is also erroneous. Corresponding Chinese forms of name have not been added to the 670 fields, so one cannot tell which characters correspond to the 6 persons. 6 of the 8 valid Chinese forms of name are represented on 22 bib records in the LC database, but there are also 4 roman script records that do not include characters. Again, the task of matching names on bib records with the authority records for the corresponding individuals would have to be done manually, against both the LC and OCLC databases. At the very least, each form of name that is represented on more than one bib record in the LC database would have to be evaluated to see whether it represents one or more persons, and then whether each person has an individual NAR. Consider how much time and resources might be required to accurately link bib records to NARs for names now represented in undifferentiated headings in the CJK languages. Currently, there are probably between 10,000 and 12,000 undifferentiated NARs for Chinese names alone. Each of these NARs currently represents at least 2 people; many represent 5 or 10 or more. In the past decade, the number of East Asian catalogers has decreased, and workloads have remained constant or increased. Therefore, I do not believe that there are sufficient resources among East Asian librarians to dedicate sufficient time to undertaking and completing the linking of authorities and bib records as recommended in the proposal. For that reason, I think that aspect of linkage is simply impractical. Thank you for considering my input. |
| 83 | we should wait for the implementation of RDA. I am fine with the current NACO practice. |
| 84 | It is good to librarians. I doubt patrons look at authority records at all. Most ILS opac do not display authority record's fields other than 1xx and 4xx. It is going to be a nightmare to the name index. |

**Q3. Comments and/or concerns regarding the method of splitting up authority records for undifferentiated personal names (See Appendix C of PCC Task Group on AACR2 & RDA Acceptable Heading Categories Final Report on p.31-33 at:** [**http://www.loc.gov/aba/pcc/rda/RDA%20Task%20Groups.html**](http://www.loc.gov/aba/pcc/rda/RDA%20Task%20Groups.html)**)**

Support:

|  |  |
| --- | --- |
| 10 | Seems OK What are plans for non-Latin-alphabet 4XX fields present in the original? This information is most important to differentiating the names and should not be lost. |
| 21 | Concern -- that this is long overdue |
| 30 | I think the step described in this report will work for authority records that only contain Latin references but not so much for those with non-Latin references. As the CJK community, we need to come up with some recommendations on how to handle the splitting of the later group. |
| 34 | Is there a program that enable machine to split an undifferentiated personal names into multiple differentiated ones? If so, split them. If not these records should remain AACR2 or mark them as "need human review". If machine cannot perform the split I will only create a new NAR for the one in hand, delete the name and 670 from the undifferentiated NAR. It will take too much time to create multiple NARs for each name listed in it. In addition, uncomfortable to create new NAR for name not in hand. |
| 35 | adding birth year |
| 52 | The proposed method sounds a good idea and a manageable method for a large file like the LCNF |
| 59 | -- In addition to "[Author of...", there are also "[Editor of...", [Joint author of..., "[Joint editor of..." etc. They all need to be taken into consideration. -- Instead of deleting the existing undifferentiated authority records right away after the records have been aplit, perhaps it would be safter to store them somewhere else until a thorough evaluation of the process has been completed. -- The process says to include "All Latin-alphabet 4XX fields preent in the original record" in each new record. I wonder whether it's possible to do something more specific? For example, perhaps the program can determine which 4XX field belongs to which new authority record based on the data in 670 fields? -- What will be done to non-Latin 4XX fields? -- It might be a good idea to identify the language of the names and involve NACO-independent catalogers to review and edit the authority records after machine conversion. |
| 65 | Who will have the time in checking existing undifferentiated names? This is a very time consuming task especially for those who do not have item(s) cited on-hand to do further investigation. |
| 68 | Use vernacular form of names in 400 fields. |
| 69 | It is a very complicated task and it cannot be split them up that easily. For example, n85054007 was populated by OCLC’s Machine-derived non-Latin script reference project based on bibliographic data, so many Li, Gang(s) are there. It is hard for machine to split them up by matching 670 fields. |
| 73 | I feel like I have been seeing more recordswith containing a non-Latin character somewhere in 670 subfield $b than six as noted in a footnote. Does this make any difference in procedure? For example, n 2012035269 and n 81138850. |
| 82 | I've seen quite a few undifferentiated name records with incorrect coding as "a" for byte 32 of the 008 field contains value “b”. This kind of records would be missed for sure by this method. |

Not Support:

|  |  |
| --- | --- |
| 16 | It does not address handling of non-Latin script references appearing in undifferentiated NARs to be split up. |
| 40 | See comments in Q 2 |
| 58 | see above Don't do it! |
| 84 | Method is ok but indexing is based on 1xx not the fixed field codes. Splitting up the record will save authority librarians some time when reviewing the record for further enhancement of the heading. |

**Q4. Comments and/or concerns regarding implication of the new policy on library systems dealing with non-uniqueness of undiff headings (See part 2 of the Discussion paper on p. 2 at:** [**http://www.loc.gov/aba/pcc/**](http://www.loc.gov/aba/pcc/) **(listed under the section PCC "What's New, Decisions, Policies and Tasks" with name: Undifferentiated Personal Names)**

Support:

|  |  |
| --- | --- |
| 10 | They will have to establish identifiers and terms that will be used to differentiate each 'entity' Perhaps data from 667 and 670 fields can also be leveraged to provide identity. |
| 21 | Comments -- no problems here at our major university. |
| 30 | The proposal of redefining "b" in 008/32 is a good resolution if it is used in conjunction with the usage of 672 proposed in the MARC discussion paper no. 2012-DP01. The usage of 672 will allow catalogers to keep track of works by the same author. |
| 34 | Need a lot of work to clear headings in local system but should be workable. |
| 52 | UNC's library system is a unicode based ILS system, the retrieval mechanism always includes non-roman field data. From the initial assessment, the split to make each heading unique will not affect retrieval but helps identify names easier. The benefit seems outweigh whatever the potential technical concerns. |
| 59 | -- Some non-Roman names become non-unique because of romanization. One possible solution for distinguishing names like this is to use the name in the original script as a qualifier. However, this scenario will only work in an Unicode environment where all scripts can be displayed simultaneously. -- Until a viable solution is in place, Undifferentiated Personal Name authority records probably should be excluded from automatic authority control to avoid future problems. |
| 65 | What happens to the libraries who are not changing to RDA due to various reasons (staff time, financial considerations)? How would the changes affecting them? |
| 69 | no opinion snice I don't know much about system functions |
| 82 | Both resolutions may work. In my opinion, it's still better to provide $c to qualify the same name heading in addition to an unique ID number. More information is better than no information. |

Not Support:

|  |  |
| --- | --- |
| 40 | Not quite relevant to what I'm doing. |
| 58 | see above |
| 84 | Adding ID number to subfield only make sense to a computer system, not patrons. If the ILS opac can define what the number and subfield code mean then it is better. Will the patrons understand what "undifferentiated" really mean? The names are the same. Again, index is big problem.. |

**Q5. Comments and/or concerns regarding implication of the new policy on PCC NACO practice (See part 3 of the Discussion paper on p. 3-4 at: at:** [**http://www.loc.gov/aba/pcc/**](http://www.loc.gov/aba/pcc/)

**(listed under the section PCC "What's New, Decisions, Policies and Tasks" with name: Undifferentiated Personal Names)**

Support:

|  |  |
| --- | --- |
| 10 | Should make it easier to create headings. |
| 34 | More work but worth the effort. |
| 59 | -- The "high-tech" environment described in Part 3 of the discussion paper is attractive. However, unfortunately, it does not exist in every part of the world yet. As metadata are being created and used globally, the issue of digital divide needs to be considered all the more. Alternative methods need to be available for systems that cannot perform at the highest level. -- Expanding the source of authority records is a good idea. However, there should still be ways to either avoid multiple authority records for the same entity being created, or link them together. |
| 82 | If eliminating a unique text string, the name authority record must have some information to identify the person. Otherwise, I doubt the value of using this kind of record. |

Not Support:

|  |  |
| --- | --- |
| 4 | In addition, we should rather allow for bibliographic records to serve as surrogate authorities, much like in cases of uniform titles for journals, unless an author is prolific or enigmatic (i.e. multiple aliases) enough to truly require an authority file. The fear amongst catalogers to not input birth or death dates in bibliographic records, even if these dates readily ascertainable from the resource itself or in tandem with other (i.e. non-DLC) national library catalogs, is a joke. Let the PCC empower others, even if these others make mistakes: PCC is not beyond error, as many of my experiences using OCLC have taught me. |
| 40 | Until the construction of name authority records has a major shift, I don't feel that the splitting would bring more benefit than the status quo. |
| 58 | see above |
| 84 | I don't think eliminating unique text string is going to work. |