

2022 ASIS&T Webinar Series December 1, 2022

“Linking Challenges for Pseudonyms (DCMI)”

webinars@asist.org

asis&t

Association for Information Science and Technology 1

Linking Challenges for Pseudonyms

Charlene Chou

Email: Charlene.Chou@nyu.edu

New York University, Division of Libraries

DCMI Webinar

December 1st, 2022

Agenda

- Background
- Scope and methodology
- Case studies
- Findings & discussions
- Recommendations
- Further research

Background

- Pseudonym: alias, fictitious name, pen names, nicknames, stage names ...
[not a true name]
- Some writers needed to use pseudonyms for **privacy** when they authored books
 - The [Brontë sisters](#) used pen names for their early work, so as not to reveal their gender and so that local residents would not know that the books related to people of the neighbourhood. The Brontës used their neighbours as inspiration for characters in many of their books.
- Using pseudonyms is indeed more important for **living writers**. For instance, a writer may write a politically-sensitive book and revealing one's real identity may be life-threatening, and thus pseudonyms are also known as **anonyms**.

Background (2)

- The Wikipedia describes pseudonyms as “part-time” names, “used only in certain context—usually adopted to **hide** an individual’s real identify.”
- In addition to individual pseudonyms, a **joint pseudonym** can be used for multiple writers writing one work together or for a “collective project such as “Nicolas Bourbaki,” adopted by mathematicians in the 1930s to stand for their collective efforts in the field of mathematics.”
- In terms of cyberspace and internet security, a **digital pseudonym** may be created to protect the privacy of users.

Background (3)

- VIAF (Virtual International Authority File) integrates many **name authority files in multiple languages** into a single platform, but it is challenging to find an author's name consistently due to various cataloging rules and vague linking.
- When an author has a pseudonym, the linking relationships become more complex or ambiguous.
- This presentation examines and explores **how to improve the linking relationships of persons with more than one identity** in VIAF through definition comparisons and case studies in terms of the semantic web.
- In light of **identity management** in cyberspace, library authority data should include the perspectives of privacy and security.

- As Hickey & Toves (2014) note, OCLC identified certain issues such as **two name authority records referring to the same entity** or a **single authority record being the mix of two entities**.
- OCLC has enhanced the VIAF data by merging related name authority data into **clusters**, assigning VIAF identifiers to the clusters and creating links between clusters as needed (e.g., between pseudonyms), but Hickey & Toves suspect that **resolving those issues “will be very difficult without deep domain knowledge.”**¹⁰ (Hickey & Toves, 2014, p. 9)
- Catalogers have been using VIAF as one of the major research resources for creating name authority records in LCNAF for several years, but found that it is imperative to use its records for persons having more than one identity with caution because its linking relationships are complex and may be misleading without domain-specific knowledge.

Scope and methodology

- Scope
 - NARs (Name authority records) of persons with more than one identity
 - Their linking relationships
- Methodology:
 - Literature review
 - Definition studies
 - Case studies
 - Comparisons
 - Definitions: personal names, identities, variant names, anonyms and pseudonyms, joint pseudonyms and partial identities, etc.
 - Finding analysis

Comparison Table of Terminologies: Pseudonyms and/or Identity Management

Term	Wikipedia	RDA	Privacy-enhanced identity management
Pseudonym	A name that a person or group assumes for a particular purpose, which can differ from their first or true name.	A name used by a person, either alone or <i>in collaboration with others*</i> , that is not the person's real name (*also called: <i>joint pseudonym</i>)	-An identifier of a subject other than one of the subject's real name -Regarding the degree of linkability, various kinds of pseudonyms may be distinguished: <i>person pseudonym, role pseudonym, relationship pseudonym, role pseudonym, role-relationship pseudonym</i> and <i>transaction pseudonym</i> -Might be an identifier for a partial identity
Alternate identity	N/A	A pseudonymous or other identity assumed by a person	Related terms: pseudonyms, partial identities, etc.
Real identity	N/A	A real person who assumes an alternate identity	The holders' civil identity; one of its real name
Person	A being that has certain capacities or attributes such as reason ...; about <i>personal identity</i> and <i>self</i> : both about what makes any particular person that particular person instead of another..."	An individual or an identity established by an individual, either alone or in collaboration with one or more other individuals.	The holder's name
Identity	Identity (social science), individuality, personal identity, social identity, and cultural identity in psychology, sociology, and philosophy	N/A	An identity is any subset of attribute values of an individual person which sufficiently identifies this individual person within any set of persons.
Individuals with more than one identity	N/A	An individual uses his or her real name as well as one or more pseudonyms (9.2.2.8)	

Comparison Table of Terminologies: Pseudonyms and/or Identity Management (2)

Term	Wikipedia	RDA	Privacy-enhanced identity management
Collective identity	The shared sense of belonging to a group	Joint pseudonym	N/A
Partial identity	N/A	N/A	An identity of an individual person may comprise many partial identities which each represents the person in a specific context or role. It is a subset of attribute values of a complete identity
Complete identity	N/A	N/A	The union of all attribute values of all identities of this person
Anonyms or anonymity	Without a name or namelessness	N/A	Anonymity of a subject means that the subject is not identifiable within a set of subjects, the anonymity set
Digital identity	Information on an entity used by computer systems to represent an external agent. That agent may be a person, organization, application	N/A	Attribution of attribute values to an individual person, which are immediately operational accessible by technical means; the identifier of a digital partial identity can be a simple e-mail address in a news group or ...
Identity management	Also known as identity and access management (IAM) is, in computer security, the security and business discipline that "enables the right individuals to access the right resources at the right times and for the right reasons".	N/A	Managing various partial identities (usually denoted by pseudonyms) of an individual person ...)

Comparisons

- Definitions:
 - Person vs. identity (RDA: real vs. [alternate identity of a person of](#))
 - “Pseudonym”: definitions vary in different languages or cultures

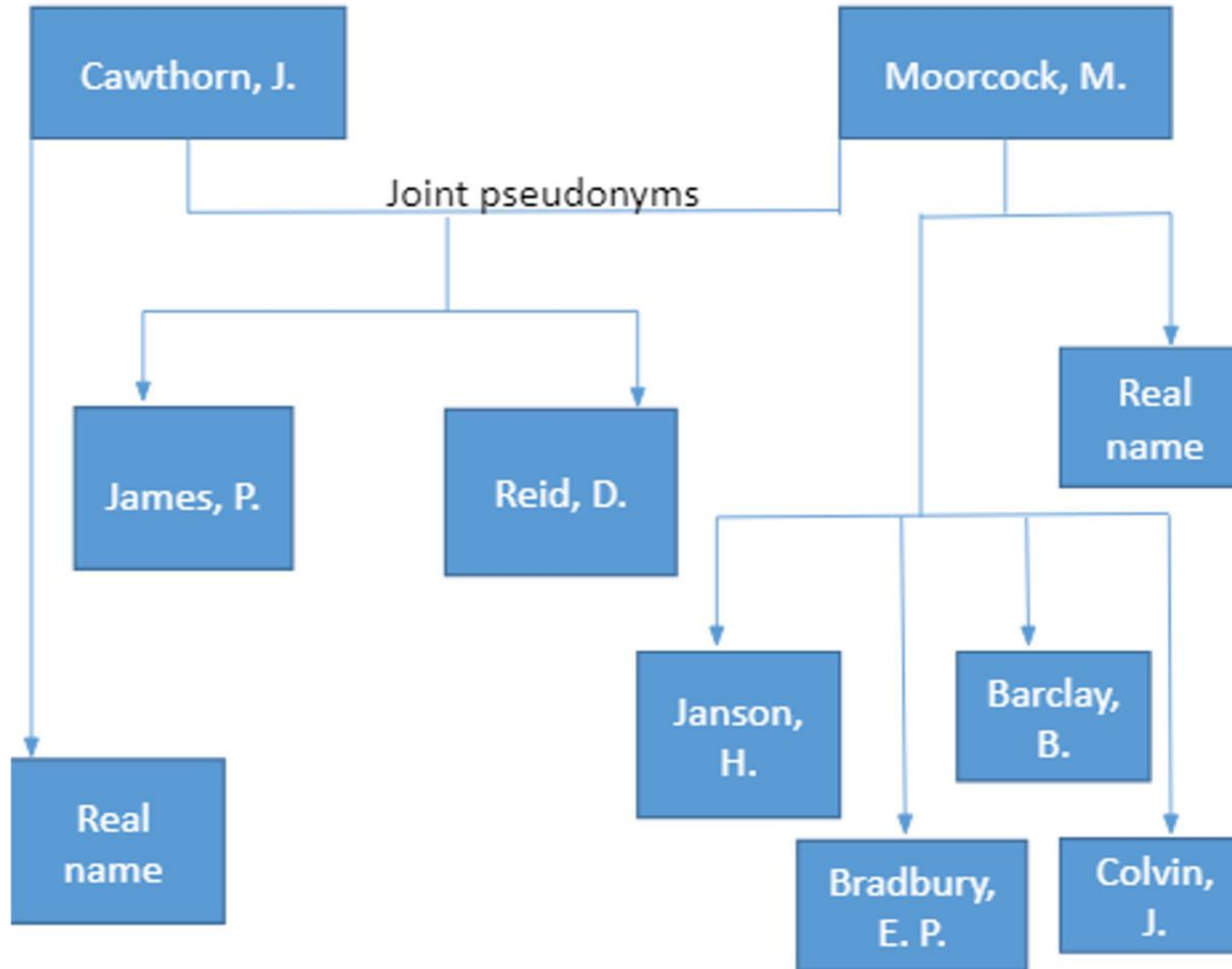
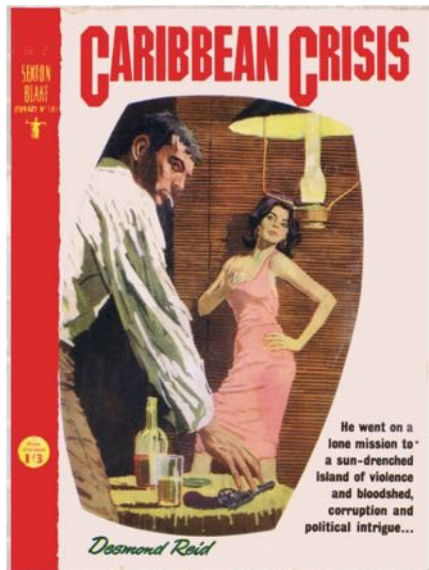
record as appropriate	Chinese	Japanese	Korean
字	zi [courtesy name]	azana [pseudonym]	cha [courtesy name]
號/号	hao [alias] or hao [pseudonym]	gō [pseudonym]	ho [pseudonym]
雅号	ya hao [esteemed name] or ya hao [nickname]	gagō [pseudonym]	

- The linking of pseudonyms and anonyms
 - LC/PCC: assumes to link the real name and one’s pseudonym(s)
 - Wikipedia: [Biographies of living persons](#): with an emphasis on privacy
- Fundamental differences in different cataloging rules
 - Original RDA + LC/PCC PS vs. other national libraries/Wikipedia, etc.
- Various types of pseudonyms and their linking issues
 - Joint pseudonym ([DCM Z1](#)) and [collective pseudonym](#) (Wikidata)

Case studies

- Mark Twain in VIAF
 - Samuel Clemens (3 groups of headings to choose + some in variants (4xx))
 - Snodgrass, Quintus Curtius: 2 groups, 4 libraries having separate records
 - Conte, Louis de: a mixture of related identities in 4xx and separate records
 - Alden, Jean Francois: no separate record in LCNAF but only two records from ISNI and Sudoc (France)
- Xiaoye, 1951 and Li, Yuan (real name in ISNI only)
- Michael Moorcock & James Cawthorn (joint pseudonyms)
- Fan, Zengxiang, 1846-1931 (with 55 pseudonyms)
 - Complex linking with mixed identity-based or person-based entities

Two authors with two joint pseudonyms and both of them published books in real names and individual pseudonyms at the same time



VIAF (Virtual International Authority File)

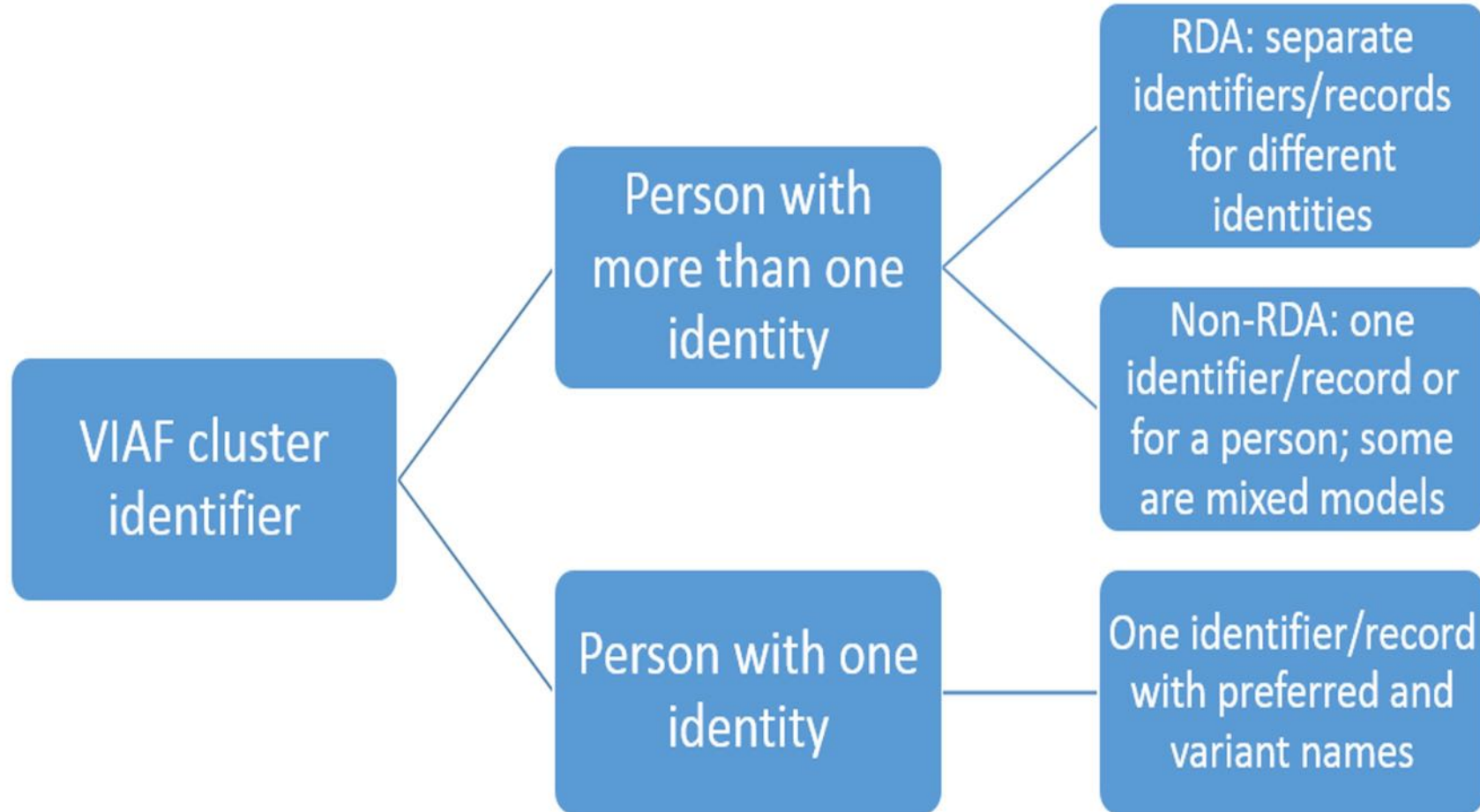
6 headings found for *Fan, Zengxiang, 1846 1931*

	Heading
1	Fan, Zengxiang, 1846-1931    樊增祥, (中國文學), 1846-1931   樊, 增祥  Zengxiang Fan 
2	Shenyunjushi, 1846-1931 
3	Fanshanjushi, 1846-1931 
4	Dongxijushi, 1846-1931 
5	Fan, Fanshan, 1846-1931 
6	Shenyunjushi 1846-1931 

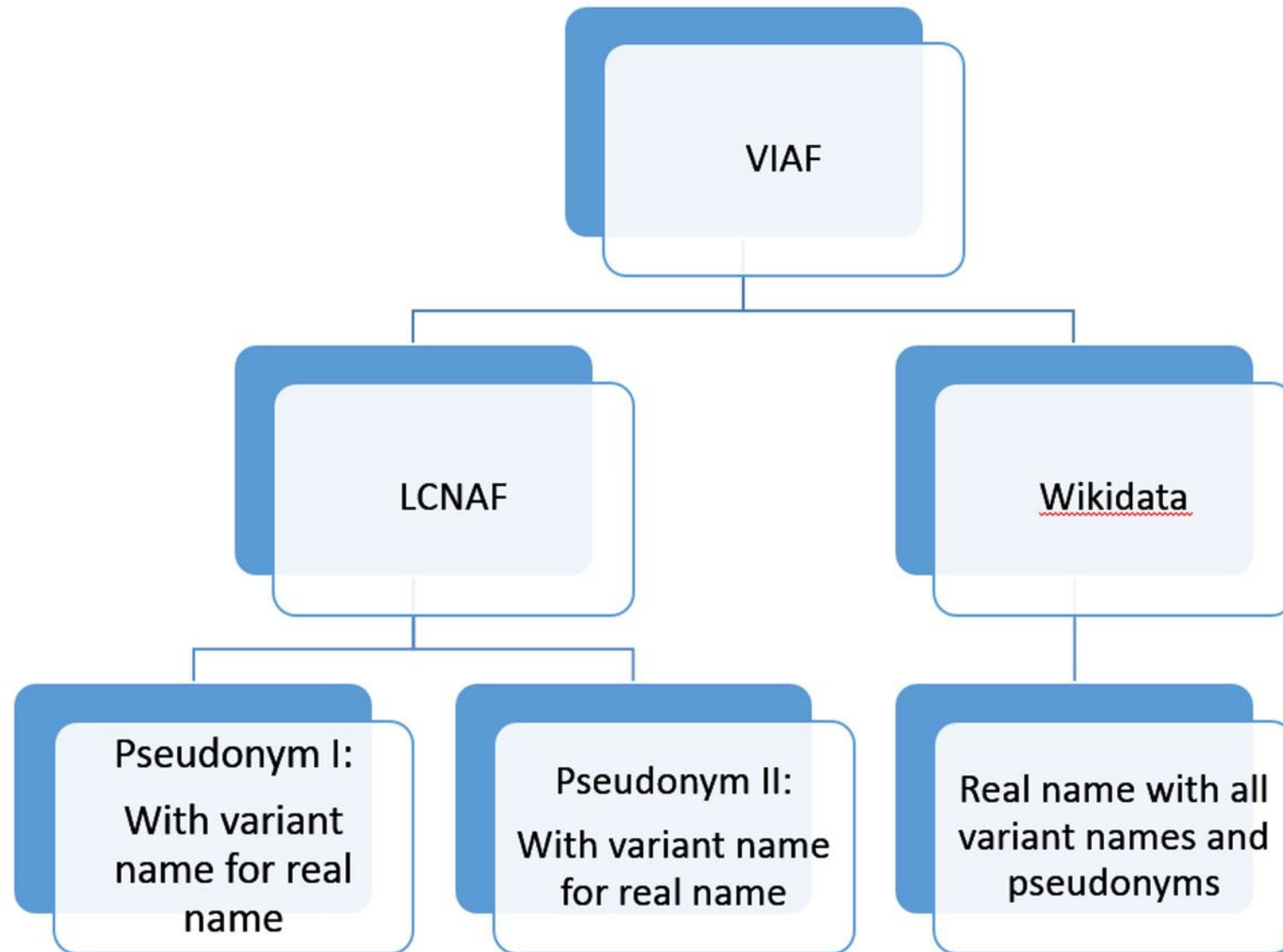
400	0	īa 東溪居士, īd 1846-1931
400	0	īa Dongxijushi, īd 1846-1931
400	0	īa 范雲老, īd 1846-1931
400	0	īa Fanyunlao, īd 1846-1931
400	0	īa 無病居士, īd 1846-1931
400	0	īa Wubingjushi, īd 1846-1931
400	0	īa 雲山, īd 1846-1931
400	0	īa Yunshan, īd 1846-1931
400	0	īa 雲門, īd 1846-1931
400	0	īa Yunmen, īd 1846-1931
400	0	īa 嘉父, īd 1846-1931
400	0	īa Jiafu, īd 1846-1931
400	0	īa 萬範水, īd 1846-1931
400	1	īa 樊山, īd 1846-1931
400	0	īa 樊山老人, īd 1846-1931
400	0	īa Fanshanlaoren, īd 1846-1931
400	0	īa 樊山居士, īd 1846-1931
400	0	īa Fanshanjushi, īd 1846-1931
400	0	īa 樊山美人, īd 1846-1931
400	0	īa Fanshanmeiren, īd 1846-1931
400	0	īa 樊方伯, īd 1846-1931
400	0	īa Fanfangbo, īd 1846-1931
400	0	īa 樊美人, īd 1846-1931
400	0	īa Fanmeiren, īd 1846-1931

Findings & discussions

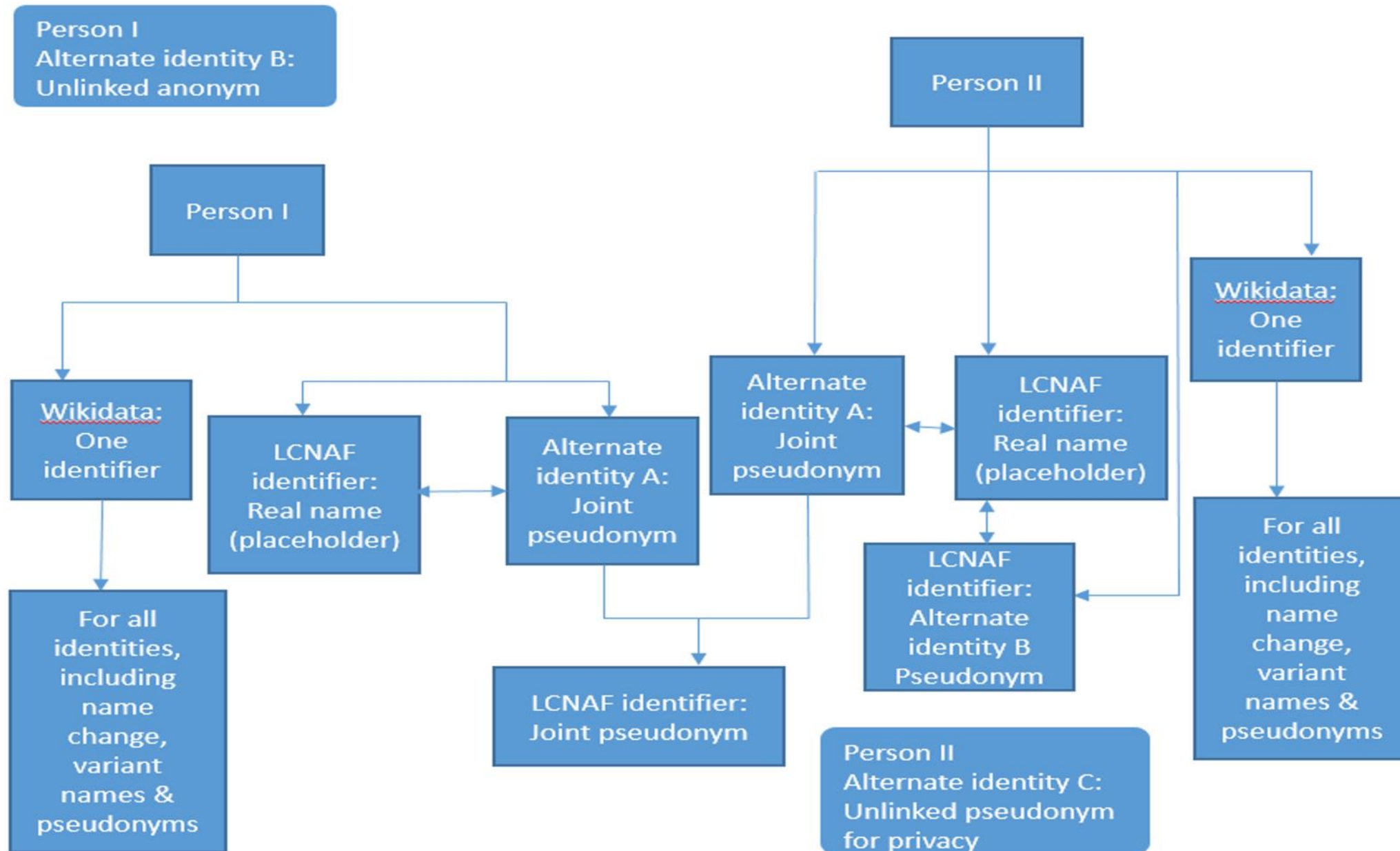
I. RDA vs. non-RDA for person with more than one identity under the VIAF cluster Identifier



A Scenario: a person with only publications under pseudonyms



Scenarios of pseudonyms and joint pseudonyms: linked, (un)linkable and orphan data



Complex issues in VIAF

- Person vs. identity-based identifier: single vs. separate record approach
- Inconsistency in linking related identities
- Inconsistency in variant names
- Diverse definitions of related names
- Pseudonyms: privacy and security issues

Comparison Table:
LCNAF, ISNI,
Wikidata and VIAF:
Basis, Labels and
Identifiers

Category	LCNAF (RDA)	ISNI	Wikidata	VIAF
Basis	Identity	Mixed identities and names	Person	Identity with linking to person-based records
Preferred name	Personal name heading (MARC tag 100)	Name	Person's name on top and in the Label column	Preferred forms
Variant names	Variant(s) (MARC tag 400)	Name	Also known as	Alternate name forms
Related identities	See also (MARC tag 500)	Related identities: (Pseud. ; Other identity, same person)—qualifier	Pseudonym	Related names (for other identities)
Related names	See also (MARC tag 663 for joint pseudonym (joint identity))	Related names: for collaboration work; organization member, etc.	Member of; influenced by;	Related names; Selected co-author, etc.
Separate record for person more than one identity	YES	YES	NO	YES, but may link to person-based record
Identifier/URI	Identity	Identity; “only assigned to clusters with a sufficient confidence rating assigned by the algorithms”*	Person (cluster)	Cluster identifier: “clusters created based on the alignments of data can change frequently”*

Recommendations

- Semantic mapping with granularities
 - RDA: Alternate identities for a variety of pseudonyms
 - Alternate identity of person & alternate identity of person of
 - PS (Policy statement) for specific instructions, e.g. [alternate identity of person](#)
 - LC/PCC practice: See the Metadata Guidance documentation: [Person elements—MG: Person: Alternate identity of person](#)
- VIAF
 - Improve VIAF linking and mapping: domain-specific experts
 - Improve its interface
 - Granularities for persons with more than one identity; group names by languages

Recommendations (2)

- LC/PCC (Library of Congress/Program for Cooperative Cataloging)
 - LCNAF (Name Authority File): use 024 (related identifiers) with caution for pseudonyms (person vs. identity)
 - [Description Cataloging Manual, Z1: Name and Series Authority Records](#)
 - Option to consider: unlink an identity per author's request (esp. for living persons)?
- Best practices needed for different languages or cultures
 - [CJK Pseudonyms Best Practices \(2019\)](#) (Council on East Asian Libraries, US)

Linking real names with pseudonyms in the name authority records (LCNAF), according to the [FAQ – LC/PCC Practice for creating NARs for Persons who use Pseudonyms](#)

Example 1.2: Person uses both the real name and one pseudonym - Two NARs; real name is added as a simple 500 see-also reference.

100	1#	\$a Sandford, John, \$d 1944 Feb. 23-
500	1#	\$a Camp, John, \$d 1944-

100	1#	\$a Sandford, John, \$d 1944 Feb. 23-
500	1#	\$w r \$i Real identity: \$a Camp, John, \$d 1944-

Reciprocal NAR is made with simple 500 see-also.

100	1#	\$a Camp, John, \$d 1944-
500	1#	\$a Sandford, John, \$d 1944 Feb. 23-

100	1#	\$a Camp, John, \$d 1944-
500	1#	\$wr \$i Alternate identity: \$a Sandford, John, \$d 1944 Feb. 23-

Description Cataloging Manual, Z1: instructions for Joint Pseudonyms

NAR for joint pseudonym:

100 1# \$a Phillips, Mark

500 1# \$w nnnc \$a Garrett, Randall

500 1# \$w nnnc \$a Janifer, Laurence M.

663 ## \$a Joint pseudonym of Randall Garrett and Laurence M. Janifer. For works of these authors written under their own names, search also under: \$b Garrett, Randell \$b Janifer, Laurence M.

NAR for first 500 on joint pseudonym NAR:

100 1# \$a Garrett, Randall

500 1# \$w nnnc \$a Phillips, Mark

663 ## \$a For works of this author written in collaboration with Laurence M. Janifer, search also under: \$b Phillips, Mark

NAR for second 500 on joint pseudonym NAR:

100 1# \$a Janifer, Laurence M.

500 1# \$w nnnc \$a Phillips, Mark

663 ## \$a For works of this author written in collaboration with Randall Garrett, search also under: \$b Phillips, Mark

CJK Pseudonyms Best Practices (2019)

Background: [Descriptive Cataloging Manual \(DCM Z1\)](#)

“Under **earlier** cataloging policies, pseudonyms used by **non-contemporary authors** (i.e., authors who **died before December 31, 1900**) were sometimes traced as **4XX references** rather than being established in a **separate NAR**. When **updating NARs for persons that contain pseudonyms in 4XX fields**, catalogers should follow current policies on pseudonyms. See the appropriate sections of Z1, including the 663 and 667 field sections. More information is available in the FAQ on personal names (<http://www.loc.gov/catdir/cps/pseud.pdf>).”

CJK Pseudonyms Best Practices (2019) (2)

- Follow [LC-PCC PS for 9.2.2.8](#) and the [FAQ on Individuals with more than one identity](#)
- Limit the work to the resource in hand
- Convert a variant in a 400 field only if there is explicit evidence of a separate identity
- No obligation to research and/or convert all 400 fields in a NAR for non-contemporary persons
- In case of doubt, leave existing 400s as variant names unchanged

CJK Q11. His NAR was treated as non-contemporary under AACR2, and the first 670 shows that he also published using his pseudonym “Fan, Fanshan (樊樊山).” Should this be treated as a separate identity?

CJK A11. Yes, with explicit evidence of separate identities, you can remove “400 1_ Fan, Fanshan” from the NAR for Fan, Zhenxiang to create a new NAR, following LC-PCC PS 9.2.2.8 and the Pseudonym FAQ. For additional pseudonyms/variant names in his VIAF record, it is okay to cite VIAF in a 670 to show that there are multiple forms in existence, but do not link to VIAF and do not add an 024 for VIAF either until best practices for the 024 field have been developed. Also you are not expected to create new authority records for every identity found in VIAF unless you have explicit evidence in hand.



LIBRARY OF CONGRESS AUTHORITIES



View this record in: [MARCXML](#) | [LC Authorities & Vocabularies](#) | [VIAF \(Virtual International Authority File\)](#)

Fan, Fanshan, 1846-1931

LC control no.	no2017104880
Descriptive conventions	rda
Personal name heading	Fan, Fanshan, 1846-1931 <i>Browse this term in LC Authorities or the LC Catalog</i>
Variant(s)	樊樊山, 1846-1931
See also	Works by this author are identified by the name used in the item. For a listing of other names used by this author, search also under Fan, Zengxiang, 1846-1931 Fan, Zengxiang, 1846-1931 <i>Browse this term in LC Authorities</i>
Associated country	China
Birth date	1846
Death date	1931
Special note	Non-Latin script reference not evaluated
Found in	Fan, Fanshan. Qin lou meng, 1914: title page (樊樊山 = Fan Fanshan)
Associated language	chi

Further research

- Smarter processing
 - Better tools + experts to refine data
- Identity management is an emerging field with complex issues
 - Cyberspace privacy and security
 - Archival domain broader than VIAF which is more driven by published items
- Goals of library linked data
 - Define and evaluate data models to make library metadata semantically linked

Development partner for OCLC Meridian (entities.oclc.org)

- Background
 - OCLC SEMI (Shared Entity Management Infrastructure), 2021: Advisory Group and tester
 - MVED (Minimum Viable Entity Description for a [WorldCat Work](#))
 - OCLC Meridian development partner, 2022-
 - [WorldCat Person](#): 7 elements and properties, referenced in MVED
- Environmental scan
 - Digital privacy & digital harassment
 - What does NYU do to offer support to ensure that someone is not harassed for doing the work we're asking them to do visibly?
 - What do we do to protect identities?
 - IT has hired a Chief Privacy Officer.
 - Research subjects
 - Faculty (Human Centered Design & Engineering) and their research team have been targeted and harassed due to their research on the election integrity partnership in 2020. ([Twitter](#))
 - Anti-racism, critical race theory, etc.

Development partner for OCLC Meridian: to test (2)

- User interface: options to consider
 - Use pseudonyms or anonyms when authors have privacy concerns.
 - An author name is removed or “does not display” per author’s request.
 - Break the linkage between real names and pseudonyms for living persons, especially per author’s request.
- Metadata and linking: per authors’ requests
 - Keep birth years but remove birth dates (security questions)
 - Replace birth years by other information such as occupation
 - Dissociate the author from one’s publication due to security concerns
- Goals of library linked data
 - Identifiers vs text strings
 - Do not rely on tons of personal information for disambiguation
 - Some faculty members felt uncomfortable with too much personal information online but emphasized the importance of credentials.

THANK YOU FOR ATTENDING



Submit a webinar proposal at
<https://www.asist.org/meetings-events/webinars/>



webinars@asist.org

A copy of the recording and a follow-up survey will be emailed within
24 hours.

asis&t

Association for Information Science and Technology

