

Report of the DAMENames Ad Hoc Committee

Introduction

In early 2018, the DAMEid group requested that Cataloging and Metadata unit examine the metadata needs for the DAME. When analyzing metadata needs in both OAKTrust and Fedora, it became clear that the lack of name authority control was causing serious problems for users, especially in the case of a single author having many entries in the author index. For example, Steven M. Wright, Royce E. Wisenbaker, Professor II in Chemical and Electrical Engineering, has 10 different entries for his name. This problem is caused by the lack of authority control and the inconsistent ways in which names are inputted into Vireo and OAKTrust. In their report to the DAMEid committee, the Metadata and Cataloging librarians strongly suggested that some type of name authority control be implemented within the DAME.

In smaller repositories with few names and fewer entities (e.g., persons, organizations, subjects, etc.), the absence of explicit disambiguation or authority control can be a manageable problem. When only a few authors share a name, it is easy to tell them apart based on the subject matter of the works attached to the name. The problem compounds as collections grow larger and the number of entities with the same name that need to be distinguished from each other increases. For example, in the large OAKTrust IR, it is hard for a user to identify the "Steven Wright" that he or she is looking for, as there are several authors so named with dozens of items in the IR. Another issue that emerges in a system with no authority control – such as OAKTrust – is that an everyday typographical error (an extra space, no period after an initial, misspellings, etc.) results in a new entry in the author list. This results in multiple names for one person and it means that there is no way for a user to easily identify all the works attributed to one author.

On June 27, 2018 Sarah Potvin presented the IDEA document for the DAMENames committee. It was requested that the committee produce a report involving the following.

- I. Review of the literature (p. 2)
- II. Description and analysis of existing name authority standards and coverage (p. 3-13)
- III. UNT and UH Name Apps (p.13-14)
- IV. Reasoning behind development of A&M Name App (p. 14)
- V. Proposed DAMENames App (p. 15-16)
- VI. Priorities as to which entities need to be included within the DAMENames system (p. 17-18)
- VII. How data is to be published and shared (p. 18)
- VIII. How the system will deal with change (p. 18-19)
- IX. Estimating time/effort needed to implement the solution, including, if applicable, time to (1) produce and (2) maintain a local authority file (p. 19-20)
- X. Next Steps (p. 20-21)
- XI. How will this solution affect legacy data/assets as well as newly-produced/added materials? (p. 21)
- XII. Conclusions (p. 22)

I. Review of the literature

Authority control as practiced within research libraries was historically weighted to encompassing monographic texts, so its innovative application to online collections has had an interesting history in the scholarly literature over the past few decades. One early piece addressing authority control in the new, broader online environment includes a bold statement: “Authority control is inherent to a catalog and without it, a file cannot be considered a catalog.” (Tillett, 1989, p. 2) While the author’s concern in this case was more on automating processes, her exhortation to collaborate—as with other productively successful shared cataloging efforts—rings true in the current milieu of linked data efforts. Once institutional repositories (IRs) had become more established in ensuing decades, the need for authority control became more and more obvious. Lucid problem statements appeared (Salo, 2009) detailing basic use cases and technological shortcomings, among other challenges, with some speculation about helpful solutions such as batch-editing and input tools with autocompletion functionality.

More recently, several projects grappling with local authority control have arisen at different institutions. A sampling from the literature includes efforts from the University of Denver (UD), the University of North Texas (UNT), the University of Houston (UH), Texas A&M University (TAMU), and a joint effort among universities in the West. UD’s tool development (Crowe & Clair, 2015) has been focused on archival applications of these methods, so not surprisingly, their path follows tools and standards developed for archives and special collections, such as those originating from the Social Networks and Archival Context (SNAC) project. UNT’s efforts (Phillips, Tarver, Hicks, & Waugh, 2013) (Waugh, Tarver, & Phillips, 2014) (Tarver & Phillips, 2016) are intended for digital collections specifically and have been manifested in their Names app (<http://digital2.library.unt.edu/name/>), which is currently in production. It features minimal metadata points per record, such as: authorized form of name, type of name, date of birth, educational and work affiliations, subject expertise, publication name variants, and the all-important assigned URI. Also in production is UH’s Cedar (Washington, Liu, & Weidner, 2018), which is a lightly customized iQvoc application that runs as a Ruby on Rails Gem. An application called Greens mints and resolves IDs for Cedar, which follows the DPLA MAP context classes and encompasses more concepts, including subjects compared to UNT’s Names, which is currently restricted to named entities, such as individuals, organizations, events, and buildings.

TAMU authority control ideation has been productive, but so far more limited to active faculty for the VIVO instance of Scholars, a Research Information Management (RIM) tool. Early plans to incorporate the University’s LDAP service with Scholars (Ilik, 2015) have been supplemented by responses to overcome some of the limitations encountered (Graham, Lee, Radio, & Tarver, 2018), including involving the individual researchers in maintaining their own profiles to keep them current.

A related initiative that also received IMLS support is a collaboration among universities in western states (*Western Name Authority File Project*, n.d.), which compiled use cases and workflows of interest from the Universities of Oregon and of Nevada, Reno, as well as Utah State. Finally, a recent white paper was especially useful, as it also confirms the importance and immediacy of the topic. This multi-institutional effort (Casalini et al., 2018) was also the tangible deliverable of an IMLS-funded project led by co-PIs Chiat Naun of Harvard and Kovari of Cornell. Its highlights include: discussion of a “minimum viable specification,” workflows with case studies of several ongoing projects, as well as typical scenarios and use cases.

II. Description and analysis of existing name authority standards and coverage

Authority control is a method used to identify entities and to differentiate between similar terms defining different entities, such as John Smith (professor) versus John Smith (English poet), Analysis (Mathematics) versus Analysis (Philosophy), or Amigos (Fraternity) versus Amigos (Musical group).

In traditional library-based practice, authority control operates by the principle of one-to-one correspondence between an entity and the defining term, with each term required to be unique in the system. In order for name authority control to work within a library database (e.g. a catalog), every name in the file should have its own unique form and correspond to exactly one “identity” (i.e., the name that a particular author may use to publish within the field of mathematics). Name information is stored in name authority records, which are linked to bibliographic records within a database.

The following authority standards in this section include approaches that draw upon both “traditional” library-based authority practices (LCNAF, VIAF), as well as more recent, emerging approaches that rely on the uniqueness of numerical identifiers rather than on the uniqueness of names represented as text strings.

It should be noted that the DAMENAME group only focused on personal names and not corporate names or subject headings. While we understand and champion the need for authority control in all cases, we recognize that the issue of personal names is the most pressing for the development of the DAME at this time. That said, our suggestions do try to take into consideration how the system proposed might be able to deal with corporate identities and with subject headings.

a. LCNAF

The Library of Congress Name Authority File, also known as the “NACO Authority File,” consists of name authority records that include authoritative data for names of persons, organizations, events, places and titles. Each entity is given an authorized text string that is designed to be unique.

A typical authority record will include: the established, unique text string, for the entity, as well as any variant (but non-authorized) forms of that particular text string (e.g., if there is an authoritative text string “Smith, Robin”, there may be a variant text string also recorded in the record, “Smith, Robyn”, if the author also sometimes spelled her name this way in her works). Records for organizations (“corporate bodies”) may include earlier or later names, as well as hierarchical structures. Other information related to the entity may also be recorded (e.g., country of birth or residence, occupation or field of activity, affiliated organizations, etc.).

The unique, authorized text string for a particular entity (e.g., Smith, Bob, 1978-) can be thought of an identifier. However, unlike many of the examples in this section, the “identifiers” used in LCNAF happen to be text-based rather than numerically-based. Since this text-based string must uniquely identify a particular entity, additional work must often be done to the form of name to make this form a truly unique identifier. For example, there may be many authors that share the text string “Smith, Bob”, so the birth date or other bits of information, such as date of birth, must often be added to help distinguish “Smith, Bob, 1978- “, for example, from other Bob Smiths that may be born in other years.

Each authority record is also assigned a number, which is incorporated within its URL. This URL can be used as a machine-readable identifier, as shown in the example below from the authority record for “Dobrovolny, Chiara Silvestri”:

LC NAF control number: no2012049144

Human readable identifier: Dobrovolny, Chiara Silvestri

Machine readable identifier: <https://lcn.loc.gov/no2012049144> or <http://id.loc.gov/authorities/names/no2012049144>

Example of a record for an author in the LCNAF

The screenshot shows the OCLC Connexion interface for an authority record. The browser title is "OCLC Connexion - [Online LC Authority File: Dobrovolny, Chiara Silvestri]". The interface includes a menu bar (File, Cataloging, Authorities, Edit, Action, Batch, View, Tools, Window, Help) and a toolbar with various icons. Below the toolbar, there are "Text Strings" and "User Tools" options. The main content area displays the authority record for ARN 9184200. The record details are as follows:

Field	Value
Rec stat	c
Type	z
Roman	1
Govt agn	1
Series	n
Ser num	n

Metadata fields:

Field	Value
Entered	20120405
Replaced	20180405073317.0
Upd status	a
Enc lvl	n
Source	c
Ref status	a
Mod rec	
Name use	a
Auth status	a
Subj	a
Subj use	a
Authref	a
Geo subd	n
Ser use	b
Name	a
Subdiv tp	n
Rules	z

Record content:

Field	Value
010	no2012049144
040	Tx #b eng #e rda #c Tx #d TxU-Mu
100	1 Dobrovolny, Chiara Silvestri
373	Texas A & M Transportation Institute #2 naf
400	1 #w nne #a Silvestri, Chiara
670	Development guidance for sign design standards, 2012 : #b t.p. (Chiara Silvestri)
670	Crash testing and evaluation of multiple mailbox supports for use with locking architectural mailboxes, 2017, PDF viewed 4 April 2018: #b t.p. Chiara Silvestri Dobrovolny, Ph.D.; associate research scientist, Texas A&M Transportation Institute) #u https://static.tti.tamu.edu/tti.tamu.edu/documents/9-1002-15-7.pdf
670	MASH evaluation of TxDOT high-mounting-height temporary work zone sign support system, 2017, PDF viewed 4 April 2018: #b t.p. (Chiara S. Dobrovolny, Ph.D.; associate research scientist, Texas A&M Transportation Institute) #u https://static.tti.tamu.edu/tti.tamu.edu/documents/9-1002-15-8.pdf

At the bottom of the interface, there are buttons for "Export", "Add", "Submit", "Replace", "Report Error", and "Validate". The status bar at the very bottom shows "25 of 50" records, "Workflow In Process", and "INS 38:56".

b. VIAF

VIAF is a service that gathers the library authority files of a large number of institutions from different countries. This includes data from national libraries and union catalogs, cultural and research institutions, as well as the entire NACO (Name Authority Cooperative file – which includes LCNAF). It has also begun to harvest data from Wikipedia. VIAF includes authority records for persons, corporate bodies, works, expressions, meetings, and geographic names. As

Angjeli, Mac Ewan and Boulet (2014) wrote, "VIAF is a major source for authority control and is becoming *the* collective reference source at the international level" (p.2).

Its objective is to share data with libraries, archives, and museums around the world, reducing costs by allowing them to reuse authority work done by others that are in multiple languages and scripts. It has an open data license (ODC-By) that lets its data to be re-used for free.

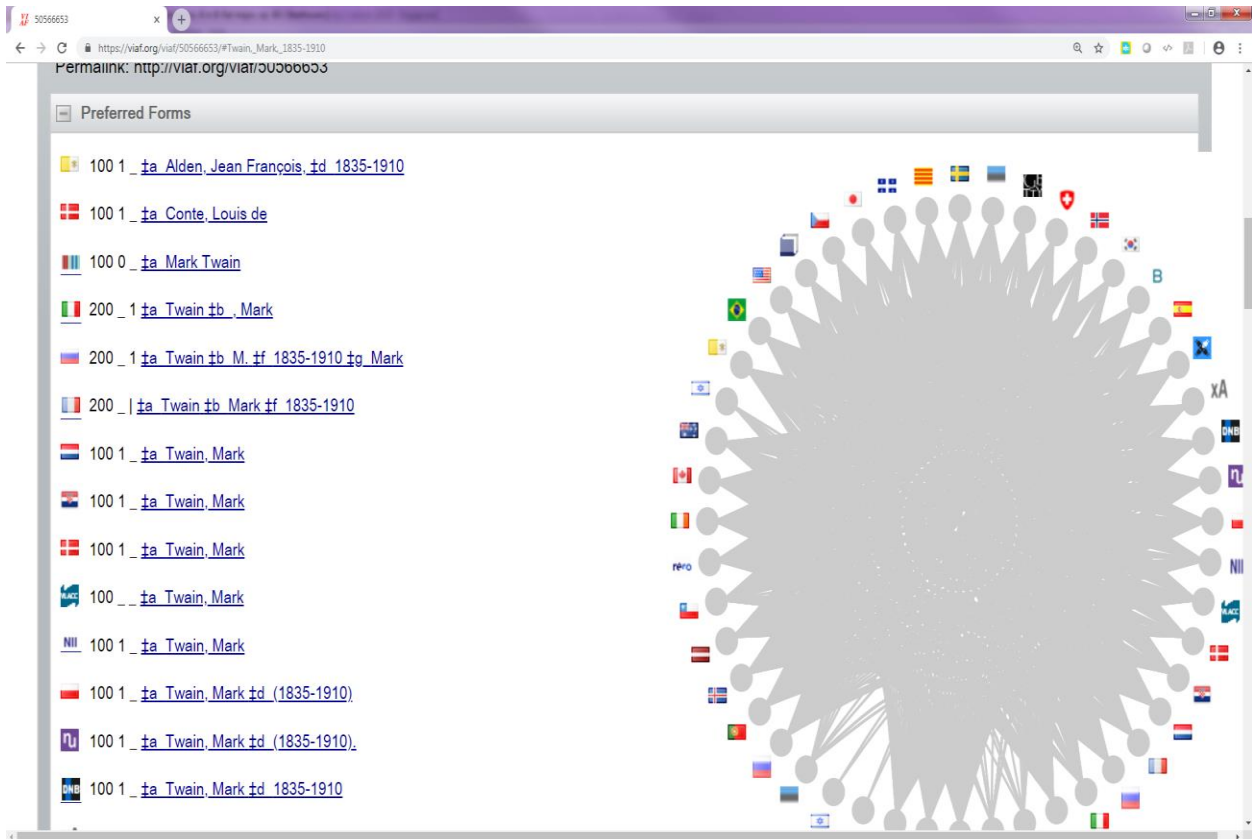
VIAF's data is organized into discrete, unique "clusters" for each separate identity. Each cluster is assigned a unique VIAF identifier. An author, for instance, may have a cluster consisting of all matching authority records from different sources (NACO, National German Library, etc.), each in the language and script of its respective country. VIAF uses specific algorithms to compare various authority records in order to determine whether they represent the same identity (e.g., "Author X" and not "Author Y") and should be assigned to the same cluster. The accuracy of the clustering often depends on the quality of the incoming authority data.

While the clusters (and their associated "VIAF IDs") are stable, the contents within a particular cluster may change over time if new authority data gets input into the system, the algorithm is run again and comes up with different results.

The VIAF ID can also be considered international in scope, as it can include authority records from multiple countries.

Example of a VIAF cluster or an author:

The screenshot shows a web browser window displaying a VIAF cluster for Mark Twain. The URL in the address bar is https://viaf.org/viaf/50566653/#Twain_Mark_1835-1910. The page content includes a header with the name 'MARK TWAIN' and a small portrait. Below the header is a horizontal row of flags representing the languages and scripts included in the cluster. The main body of the page lists various forms of the author's name and dates, such as 'Twain, Mark, 1835-1910', 'Twain, Mark', 'Твен, М. 1835-1910 Марк', 'Mark Twain', 'توين, مارک, 1910-1835', 'טוין, מרק, 1835-1910', 'Твен, Марк, 1835-1910', 'Twain, Mark, pseud.', 'Twain, Mark, pseud., 1835-1910', '1910-1835, توین, مارک', 'Twain, M. 1835-1910 Mark', 'Twain, Mark (American humorist, novelist, and artist, 1835-1910)', 'Conte, Louis de', 'Твен, Марк, псевд., 1835-1910', and 'Alden, Jean François, 1835-1910'. At the bottom, there is a 'Preferred Forms' section with the entry '100 1 _|a Alden, Jean François, |d 1835-1910'. The browser's taskbar at the bottom shows several application icons.



c. **ORCID**

ORCID is a global, unique, and persistent identifier targeted for researchers. Unlike ISNI, researchers or their institutions must initiate the assignment of their own ORCID IDs by registering with the service. Also unlike ISNI, researchers input metadata about themselves in their ORCID records to create online profiles and maintain control over their content.

Like ISNI, ORCID is a nonprofit organization with its own governing body and community, but unlike ISNI it is not an ISO standard. It is also similar to ISNI in that it is designed to serve as a stable identifier across different domains, geographic locations, and platforms, and has the ability to be linked to other identifiers, such as ISNI, ResearcherID, etc.

Its structure (a 16-digit number) is also compatible with ISNI's structure. ORCID IDs are randomly assigned by the ORCID Registry from a block of numbers that do not conflict with ISNI-formatted numbers.

While ORCID IDs have traditionally only been created for people, they will also be able to be assigned to organizations in the future. In January 2017, a working group (<https://orcid.org/content/organization-identifier-working-group>) explored how to disambiguate organizations that researchers are affiliated with. There are now plans to establish a registry of ORCID identifiers for organizations.

Example of an ORCID web page for a researcher

Robert (0000-0003-2769-1270) x

ORCID uses cookies to improve your experience and to help us understand how you use our websites. [Learn more about how we use cookies.](#) Dismiss

Search English

ORCID Connecting Research and Researchers

5,494,159 ORCID iDs and counting. [See more...](#)

Robert

ORCID ID
<https://orcid.org/0000-0003-2769-1270>
 Print view

Employment (2) Sort

- Kitsap County Sheriff's Office: Port Orchard, Washington**
 2010 to 2017 | Volunteer Chaplain (Sheriff's Office)
 Employment
 Source: Robert Preferred source
- United States Air Force: Washington, D.C.**
 1983 to 2007 | Wing Chaplain
 Employment
 Source: Robert Preferred source

Education and qualifications (8) Sort

- Concordia Seminary: St. Louis, Missouri**
 2013 to 2016 | Doctor of Ministry (D.Min.) (Military Chaplaincy)
 Education
 Source: Robert Preferred source
- USAF Air War College: Maxwell AFB, Alabama**
 2003 to 2004 | Diploma (Military History & Military Strategy)

Help

d. **ISNI**

ISNI (International Standard Name Identifier) is a global, ISO-certified standard for the “unique identification of the public identities of people and organizations across all fields of creative activity” (Smith-Yoshimura et al., 2016, p. 3). One ISNI is assigned per public identity of a person or organization across all activities and roles it may play with regard to creative contents. It has several objectives: 1) To serve as a platform-independent “bridge” identifier across different communities (e.g., scientific communities, publishers, rights organizations, libraries and museums, etc.); 2) to serve as a stable and unique identifier across different domains and geographic locations; 3) to act as an interoperable identifier by mapping to other types of identifiers, both international (e.g., ORCID) and local.

The scope of ISNI is a lot broader than ORCID. In addition to researchers, they can also be assigned to other kinds of authors, performers, artists, etc. that create content, but are not affiliated with academic or research-based institutions or organizations.

Unlike ORCID, ISNIs are not created by the individuals they are assigned to. Instead, they are either created manually within the ISNI database by individuals who are authorized by ISNI’s governing body to do so, or the data is harvested using automated means from external databases. The ISNI database harvests data from a variety of sources, including VIAF, a global source of authority data, as well as publishers and vendors.

The individual who is assigned an ISNI has little to no control over the content of his or her record within the database and cannot edit it, unlike ORCID or ResearcherID. However, records

in the ISNI database are viewable by the public, who may request changes to the ISNI Quality Team.

Example of an ISNI number: 0000 0004 6436 3572

It can also be in the form of a URI: <http://isni.org/isni/0000000464363572>

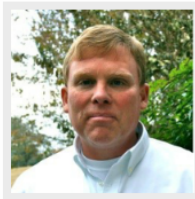
Example of a record in ISNI for an author:

The screenshot shows the ISNI Search website interface. At the top, there is a navigation bar with links for Search, Results, Advanced, Terms of Use & Use of Cookies, Help, and Privacy. The main search area includes a search box with the text 'blackstock, tiffany', a search button, and a 'less' link. Below the search box, there are tabs for 'shortlist', 'title data', and 'search history'. The search results section shows 'results search [and] (Any Phrase (APH:)) blackstock, tiffany | 1 hits'. On the left side of the record, there is a yellow box with the text: 'Please help us improve this record. If you have any supplemental information about the identity listed here, please click in this box to go to the contribution form. Thank you in advance! authority'. The main record details are as follows:

ISNI:	0000 0004 6436 3572
Name:	Blackstock, Tiffany Blackstock, Tiffany L.
Creation class:	text
Creation role:	author editor
Related names:	Stevens, Charles R. (co-author) Texas A and M University School of Law Texas Transportation Institute
Titles:	Demonstration of unmanned aircraft systems use for traffic incident management (UAS-TIM) Texas A&M Law Review
Notes:	JD Candidate https://static.tti.tamu.edu/tti.tamu.edu/documents/PRC-17-69-F.pdf
Sources:	PCC

e. **Scholars@TAMU**

Scholars@TAMU, developed by the University Libraries using the VIVO open collaborative software, is a profile system that hosts searchable expertise for faculty and TAMU organizations. Scholars@TAMU gathers data from institutions-level/enterprise systems, publicly available research data (e.g., grants and publications), and other authoritative sources. The data is compiled into a profile that faculty can edit to best represent their scholarship and expertise. Since this system is locally developed, the University Libraries has authority to use, modify, or update its data and database. Scholars@TAMU meets all 12 criteria that this committee think appropriate to be considered as standard for identification systems.



Herbert, Bruce | Professor

Positions

- ▶ Director, Office of Scholarly Communications and Professor, [University Libraries, Texas A&M University](#)
- ▶ Professor, [Geology and Geophysics, College of Geosciences](#)

[Co-author Network](#)

[Map of Science](#)

[Co-investigator Network](#)

Contact Info

beherbert@tamu.edu
 1 979 845 1083

Websites

- ▶ [Google Scholar](#)
- ▶ [Library Profile](#)
- ▶ [Mendeley profile](#)
- ▶ [My Research Homepage](#)
- ▶ [Scholar Plot of BE Herbert](#)

I currently serve as the Director of the Office of Scholarly Communications in the Sterling C. Evans library At Texas A&M University. As Director of OSC, I am responsible for strengthening the Library's efforts in scholarly communications and open access through engagement and collaboration with the faculty across campus at Texas A&M University. In addition, we are developing library services that support interdisciplinary research teams and enhance research translation and the societal i (... [more](#))

Research Areas

[Academic-industrial collaboration](#) | [Communication in learning and scholarship](#) | [Open access](#) | [Research--Management](#) | [Soils--Quality](#)

ORCID

[0000-0002-6736-1148](#)

- Academic Background
- Publications
- Grants/awards
- Recent Courses
- Works By Students
- Contact
- View All

Academic Background

f. **ResearcherID:**

ResearcherID is a unique identifier that is assigned to a researcher that registers for one on the ResearcherID.com website. It is designed to be persistent and to let researchers manage information about their publications, identify potential collaborators on the website, and differentiate their identities from other researchers who may have identical names. According to its website, ResearcherID is also “ORCID compliant” and integrates with Web of Science, “allowing you to claim and showcase your publications from a single one account.” Thus, it is similar to ORCID, in that it lets researchers freely self-register on the website and build their own online profiles.

Herbert, Bruce E [Return to Search Page](#) [Get A Badge](#) [ResearcherID Labs](#)

ResearcherID: L-2170-2015 My Institutions (more details)

Other Names: B. Herbert; B.E. Herbert; Herbert, Bruce E. Primary Institution: Texas A&M University

URL: <http://www.researcherid.com/rid/L-2170-2015> Sub-org/Dept: University Libraries, Geology & Geophysics

ORCID: <http://orcid.org/0000-0002-6736-1148> Role: Faculty

My Publications

My Publications (82) [View Publications](#) [Citation Metrics](#)

ResearcherID Labs [Create A Badge](#) [Collaboration Network](#) [Cling Articles Network](#)

My Publications: View

This list contains papers that I have authored.

62 publication(s) Sort by: Publication Year Results per page: 10

- Title: A Flow Adsorption Microcalorimetry-Logistic Modeling Approach for Assessing Heterogeneity of Bronsted-Type Surfaces: Application to Pyrogenic Organic Materials
Author(s): Harvey, O. R.; Leonce, B. C.; Herbert, B. E.
Source: Environmental Science & Technology Volume: 52 Issue: 11 Pages: 6167-6176 Published: 2018
Times Cited: 1
DOI: 10.1021/acs.est.8020104
- Title: Physicochemical Characterization of Binder-Aggregate Adhesion Varying with Temperature and Moisture
Author(s): Cucalon, L. G.; Bhamin, A.; Kassem, E., et al.
Source: Journal of Transportation Engineering Part B-Pavements Volume: 143 Issue: 3 Published: 2017
Times Cited: 0
DOI: 10.1061/(asce)1095-2403(2017)
- Title: An index-based approach to assessing recalcitrance and soil carbon sequestration potential of engineered black carbons (Biochars)
Author(s): Harvey, O. R.; Kuo, L. J.; Zimmerman, A., et al.
Source: Environmental Science & Technology Volume: 46 Pages: 1415-1421 Published: 2012
Times Cited: 127
DOI: 10.1021/ee2040396
- Title: Generalized two-dimensional perturbation correlation infrared spectroscopy (2D-PCIS) reveals mechanisms for the development of surface charge and recalcitrance in plant-derived biochars
Author(s): Harvey, O. R.; Herbert, B. E.; Kuo, L. J., et al.
Source: Environ. Sci. & Technol. Published: 2012
Times Cited: 63
DOI: 10.1021/ee302971d

g. Scopus Author ID

Scopus is a subscription-based, multidisciplinary database of peer-reviewed journal articles, books, conference publication, and other literature. Authors with publications indexed in Scopus are automatically assigned a Scopus Author ID. Users can search the lookup tool to locate an author's profile, which includes the identifier, references, citations of work, h-index, and subject areas. Scopus Author ID meets all four required criteria by the policy. However since the ID is a product of Elsevier, it does not fulfil the criteria related to open and non-profit.

Scopus Search Sources Alerts Lists Help Dong Joon Lee

Author details About Scopus Author Identifier

The Scopus Author Identifier assigns a unique number to groups of documents written by the same author via an algorithm that matches authorship based on a certain criteria. If a document cannot be confidently matched with an author identifier, it is grouped separately. In this case, you may see more than one entry for the same author.

< Return to search results 1 of 2 Next > Print Email

Herbert, Bruce Eric [Follow this Author](#) h-index: 20 [View h-graph](#)

Texas A and M University, Office of Scholarly Communications, College Station, United States
Author ID: 7007123830 View potential author matches

Other name formats: [Herbert, B.](#) [Herbert, Bruce E.](#) [Herbert, B. E.](#) [Herbert, Bruce](#)

Subject area: [Environmental Science](#) [Earth and Planetary Sciences](#) [Chemistry](#) [Engineering](#) [Chemical Engineering](#) [Social Sciences](#) [Agricultural and Biological Sciences](#) [Physics and Astronomy](#)
[Pharmacology, Toxicology and Pharmaceutics](#) [Materials Science](#) [Energy](#)

Documents by author: 57 [Analyze author output](#)

h. Wikidata

Wikidata acts as central storage for the structured data of its Wikimedia sister projects including Wikipedia. All people can add and edit data, and use it for free. Wikidata is a document-oriented database, focused on items. Each item represents a topic and is identified by a unique number, prefixed with the letter Q (known as a "QID"). For example, the item for the topic Benjamin Franklin is Q34969. Wikidata meets three required criteria: uniqueness, compatibility, and actionability. However, it does not guarantee persistence or security due to Wikimedia's mission: everyone can add and edit data.

Albert Einstein (Q937)

German-born physicist and founder of the theory of relativity
 Einstein | Einstein, Albert

▼ [In more languages](#) Configure

Language	Label	Description	Also known as
English	Albert Einstein	German-born physicist and founder of the theory of relativity	Einstein Einstein, Albert
Korean	알베르트 아인슈타인	독일과 미국의 물리학자	아인슈타인 알버트 아인슈타인
Spanish	Albert Einstein	físico germano-estadounidense, autor de la teoría de la relatividad	Einstein
Traditional Chinese	阿爾伯特·愛因斯坦	德裔美國物理學家，相對論創立者	愛因斯坦 阿爾貝特·愛因斯坦 愛恩斯坦 艾爾伯特·愛因斯坦 愛爾伯特·愛因斯坦 艾因斯坦 愛恩思坦 愛因思坦 亞伯特·愛因斯坦

[All entered languages](#)

Statements

instance of	Q5 human
	2 references

image	Q5 
-------	---

i. *Google Scholar*

Google Scholar is a search engine that indexes the metadata or full text of scholarly articles. It is run by Google with free public access. Google Scholar searches across a wide range of scholarly literature, including articles, books, theses, conference papers, and technical reports. Sources include academic publishers, professional societies, university repositories, and other scholarly websites. Google Scholar meets three required criteria: uniqueness, compatibility, and actionability. It does not guarantee persistence because users of Google Scholar can change their privacy setting from public to private. Google also has a strict data use policy which limits its API use.



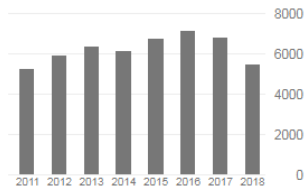
Albert Einstein

Institute of Advanced Studies, Princeton
 No verified email
 Physics

FOLLOW

Cited by VIEW ALL

	All	Since 2013
Citations	122368	38717
h-index	113	68
i10-index	370	223



TITLE	CITED BY	YEAR
Can quantum-mechanical description of physical reality be considered complete? A Einstein, B Podolsky, N Rosen Physical review 47 (10), 777	17283	1935
Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt A Einstein Ann. Phys. 17, 132-148	11161 *	1905
On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat A Einstein Annalen der Physik 17, 549-560	9599 *	1905
Zur Elektrodynamik bewegter Körper A Einstein	5411 *	
Sitzungsber. K A Einstein Preuss. Akad. Wiss., Phys. Math. Kl 3, 18	4830 *	1925

j. **Microsoft Academic**

Microsoft Academic developed by Microsoft Research is a public web search engine for scholarly literature. It provides free public API access and supports linked data technology. Microsoft Academic meets all criteria except having a gatekeeper, persistence, security, and being non-profit. Since Microsoft is a for-profit company, the system and service always include risks to be closed. Compared to its competitor, Google Scholar, Microsoft Academic seems to have smaller user group.

Microsoft Academic

Sign in or Sign up

Albert Einstein
Institute for Advanced Study

474 PAPERS 94,973 CITATIONS*

FOLLOW

CLAIM

Top Publications

[\[book\] On the Electrodynamics of Moving Bodies](#)

2014, Albert Einstein

Fields of study: classical mechanics, physics

Citations (1,337) * Download Share Cite

[Die Feldgleichungen der Gravitation](#)

2006, A. Einstein

Fields of study: einstein, gravitation, theoretical physics, classical mechanics, physics

Citations (374) * Source Share Cite

[Quantentheorie des einatomigen idealen Gases](#)

2006, A. Einstein

Fields of study: einstein, mathematical physics, physics

Citations (2,885) * Source Share Cite

Co-authors

Zugangsinformation

- John Stachel
- Robert Schulmann
- Martin J. Klein
- Diana Kormos Buchwald

Show More ▾

Co-author Affiliations

- Franklin Institute
- Institute for Advanced Study

Research Topics

- Physics
- Classical mechanics
- Einstein
- Mathematical physics

k. **Dimensions**

Digital Science (London, United Kingdom) launched Dimensions in January 2018. Dimensions is a scholarly search engine that provides free public access, although it has commercial version with much more data access including grants, patents, and clinical trials. Dimensions meets three required criteria: uniqueness, compatibility, and actionability, but it as many other identifier systems from for-profit companies, does not guarantee persistence and secureness of its data.

The screenshot displays the Dimensions website interface for a researcher profile. On the left, there are filter categories such as 'PUBLICATION YEAR', 'RESEARCHER', 'FIELDS OF RESEARCH', 'PUBLICATION TYPE', 'SOURCE TITLE', 'JOURNAL LIST', and 'OPEN ACCESS'. The central part of the page shows the profile for 'M Arul Kumar' from 'Los Alamos National Laboratory - Los Alamos, United States'. It includes an 'Overview' section with 19 publications and 239 citations, and a 'PUBLICATIONS' section listing two articles. On the right, the 'ANALYTICAL VIEWS' section shows 'FIELDS OF RESEARCH' with a list of categories and counts, and an 'OVERVIEW' section with a line graph showing 'RCR Mean' (0.28) and 'FCR Mean' (3.82) from 2009 to 2018.

III. **UNT and UH Name Apps**

- University of North Texas: “UNT Names” <http://digital2.library.unt.edu/name/>
UNT Names notes that it “is for reference, disambiguation and storage of name records,” specifically “names used by the Libraries in its various digital library systems and collections.” Names is a Django (Python web-stack) application by Mark Phillips.
The five categories that Names covers are:

- Personal names (8486 entries)
- Organizational names (976)
- Event names (7)
- Software names (3)
- Building names (20)

There were many qualities about UNT Names that the committee liked: the name disambiguation, a unique URI for each entity, biographical information that includes affiliations and subject areas, external links (such as LinkedIn). However, there are things we would like for our uses that are not included: external links to other name resources such as ISNI, VIAF, etc., linkage back to items in the repository, and information as to where/why the name is included in the app.

- University of Houston: “Cedar”
<https://vocab.lib.uh.edu/en.html>

Compared to UNT Names, UH’s Cedar encompasses a much more ambitious vocabulary range, based on the iQvoc semantic web vocabulary system (<http://iqvoc.net/>), which is also a Ruby on Rails product “originally created and is being maintained by innoQ Deutschland GmbH.”

Cedar’s broad categories include:

- Agent (including both Organization [304] and Person [4182])
- Collection (5)
- Concept
- Place
- Time period (4)

The use of Cedar as a name authority entails a commitment to a hierarchy of classes to characterize the entities being described. Deployment of Cedar also entails the deployment of UH's ARK minting app Greens. Both of these applications are written in Ruby on Rails.

IV. Reasoning behind development of A&M Name App

The committee looked at both the UNT and UH name apps, but did not find either solution attractive. The UH Cedar and Greens apps are written in Ruby on Rails, and UNT Names app is written in Django (Python). This would complicate a local deployment and additional feature development. Cedar also involves unnecessary ontological commitment to a class hierarchy describing the entities it identifies. While many name authority apps involve commitment to classes or types for entities, we consider this an unnecessary expansion in the scope of the proposed app. The tool will be much simplified by avoiding the use of types. Cedar and UNT Names app supports aliases and alternate URIs (and thus meets lots of our use cases), but it does not support lazy fetching of data from those alternate URIs to enrich the information displayed in the app. The Committee believes that the proposed name app needs to pull from other sources and work well with SCHOLARS@TAMU and OAKTrust

The DAME is very much an ecosystem which needs all component parts to work together. The committee recommends that in order to do that, a name authority app should be developed specifically for A&M's needs

We have a need for resolution and redirection services in our app – in addition to providing an authority for author names, the development team has a requirement for UUIDs of images and other resources in the DAME. In this way, images and other resources can be looked up by UUID and resolved to their URLs which may change over time. Also, as the app becomes established and new use cases arise, there will be considerable custom work to grab data on the fly from LDAP, VIVO, VIAF, ORCID, and the IRs, and such custom work would not go smoothly for our shop in a Django or Rails stack, especially if the original code base did not have such concerns in mind at the outset of its development. Finally, the committee believes that having linked data within the App that harvests information from established name authority resources will help integrate the A&M Dame within a large IR community.

V. Proposed DAMName App

The proposed DAMNames app will unitize entities (people) found in LDAP as a basis for the initial loading of names. This will be of importance for SCHOLARS@tamU and for the Faculty Publications in OAKtrust. The DAMName App will include an ID minter which has tentatively been named EIDer. The goal is to create the most efficient identifier for each entity utilizing the minimum amount of data necessary for disambiguation. It will also include a resolver that can be made to work between EIDer and the DAME repositories (DSapce, Fedora, Avalon, etc.)

The DAMName would work by creating a UUID which would then be attached to a namespace. The program will look at LDAP, VIVO (SCHOLARS@tamU), and chosen outside name resources (such as ISNI, LCNAF, VIAF, etc). If there are no matches, then a proposed ID will be created in a local name authority database. The ID will be minted once it is verified that entity matches a name found in the DAME.

One of the most important elements of the DAMName app depends on LDAP. LDAP means "Lightweight Directory Access Protocol". It is a widely-accepted standard for interacting with directory information over a network. The division of Information Technology at Texas A&M University operates an LDAP server that makes available directory information on personnel, affiliates, and organizations related to the university. Public information in the TAMU LDAP can be accessed via a RESTful API: <https://mqs.tamu.edu/rest/docs/>. Campus service providers that have an application specific need for additional non-public resources and agree to adhere to appropriate policy can be granted access – see http://infrastructure.tamu.edu/identity/access/access.html#tab_tab1.

Another important element are the external name resources. The committee compared the various name resources – both internal and external – and rated them based on an identifier quality criteria¹. Although we did not use all the criteria, we did focus on the characteristics of: uniqueness, persistence, compatibility and resolvability.

Uniqueness is important because each identifier must have a one to one correspondence between the person and the identifier. It also facilitates collocation

Persistence is important because we don't want these things disappearing on us.. The data must be reliable and stable over time. Otherwise there will be too many broken links. Persistence is stable uniqueness over time.

Compatibility is the ability to use with the main internet naming schemes (URL or URN). Essential for Linked Data.

Action ability/Resolvability is the ability to locate the object using an identifier string. Enables collocation. Ensures that correct information is retrieved. Provides confidence in the system. Identifies flaws within the system that may require human intervention to correct.

¹ Lee, D. J., & Stvilia, B. (2014). Developing a data identifier taxonomy. *Cataloging & Classification Quarterly*. 52(3), 303-336. doi:10.1080/01639374.2014.880166

	TAMU LDAP	ISNI	ORCID	VIAF	LCNAF	Scholars@ TAMU	Researcher ID	SCOPUS ID	WIKI-Data	Google Scholar	MS Academic	Dimensions
Uniqueness	1	1	1	1	1	1	1	1	1	1	1	1
Controlling body (Gatekeepers)	1	0.5	0.5	1	1	1	0	1	0	0	0	0
Persistence	0	1	1		1	1	1	1	0	?	?	?
Verifiability	1	1	1	1	1	1	1	1	1	1	1	1
Compatibility	1	1	1	1	1	1	1	1	1	1	1	1
Interoperability	0	1	1	1	1	1	1	1	1	1	1	1
Actionability/Resolvability	Local	1	1	1	1	1	1	1	1	1	1	1
API provided?			1	1	?	1	0	1	1	0	1	1
Support RDF	0	1	?	1	1	1	?	1	1	0	1	?
Security	1	1	1	1	1	1	1	1	0	0	?	?
don't pay to get information?	1	1	1	1	1	1	1	0	1	1	1	1
Non-profit	1	1	1	1	1	1	0	0	1	0	0	0
TOTAL	7	10.5	10.5	12	11	12	8	10	9	6	8	7

0=Feature is absent; 1=Feature is present

This relates to personal names, not corporate or subject.

The above recommendations are based on the current available external identifiers. Please be aware that other and newly created external identifier sources may emerge which meet the required criteria and therefore should be considered for inclusion within the EIDer system.

VI. Priorities as to what elements need to be included within the DAMENAME system

With the development of a name app, certain criteria need to be met in order to ensure fulfillment of our local use cases and smooth operation with related systems. There are two categories - the criteria for each individual records, which will identify one person/entity, and the criteria for the system as a whole.

Criteria for the individual records

- Unique identifier
 - There will be one identifier per entity. This will ensure that name disambiguation is taking place and that the user can be assured that the one identifier identifies the person they are seeking.
 - Will not be human readable
- Canonical human readable name
 - This may not be unique across the whole database
 - However, in combination of the unique identifier it will ensure that the correct entity is linked to the right entry
 - It must be human readable to assist in human catalogers in identifying which identifier needs to be used
- List of name variants (i.e., aliases) with optional time period of the use of the variant and descriptive notes
 - This includes name variations that are found on works within the DAME
 - The time period will take care of identifying name changes as well as help identify when the author was active
 - Variant names should be human readable so that users may use them to search for potential names of the person being sought
 - Notes could mention the source of the name, justification for the time period, or other details
- List of external URIs and the type (ex. LDAP, VIVO, etc.)
 - LDAP is the core resource for identifying works by current A&M faculty
 - Other resources such as VIAF and ISNI provides authority control and helps to link the author to resources outside of the DAME
- Ability to add or delete metadata on each individual record (both manually and automatically)
 - To enable editing and updating of records, as well as to assist in clean up
- Pull university affiliation and directory information from LDAP on demand for display purposes
- Logging for all change events, including time and username

Criteria for the system

- Ability to look up either by URI or human readable name (including canonical or non-canonical variations of the name)
- Ability to add or delete external name authorities types (e.g. If ISNI becomes defunct or if Google Scholar no longer annoys, add it)
- For names not found in external name authorities, in the future we could reference a locally created knowledge base

- Minting new identifiers/records for names not found
- Notify MaC about needing to create name in local knowledge base.
- Logging for all change events, including time and username
- Utilize UTF-8 for all text and API interactions
- Custom, individual modules to fetch information based on name authority type of a entities URI. This will be done only for LDAP in version 1.0
- Need some type of facility for merging entities
- Assistance for record reconciliation (e.g., notification of possible duplicate entities when new one is created)

Other considerations

- For historical names (ag extension, etc.) not found in existing external name authorities, provide an external knowledge base

VI. How data is to be published and shared

The committee was asked to answer the following questions: **“If a name app is recommended: how will the data be published and shared? Is data publication a goal? Or do you recommend limiting access to the Libraries?”**

Yes, data publication is an important goal. The committee has agreed that the data should be openly accessible to all and shareable both within the A&M system and with other libraries. Utilizing linked data and pulling information from outside name resources such as ISNI, VIAF, etc., means the data with the DAMName App will be shared. However, the public interface for the app has not yet been discussed and how data will be published depends largely on the lesson's learned from developing a prototype in Phase II

We do not recommend limiting access to only the libraries for the following reasons

- We want to encourage open access in scholarship
- Publishing the data will help promote A&M scholars and research in the wider academic communities
- Due to the integration and utilization of outside name resources, it would be feasible to make the data accessible to all

VII. How the system will deal with change?

One extremely important criterion that any name app must deal with is change. Personal names change less than corporate names, but there are still variations and name changes that occur with life events, such as marriage and divorce. There is also the important fact that any document may represent a name differently, such as: using initials instead of the first or middle name, hyphenating last names (or not),

including a middle name (or not), using an initial for a middle name, using a nickname as opposed to the legal name, etc.

Any name app would have to deal with the above very common scenarios and others that are not so common. The committee will deal with the specifics of how to handle these changes in more detail during the second phase of the DAMName group, but some suggestions and issues that we have discussed are:

- Put a ward at VIREO which will assist in mapping individual entities (i.e. people) to the approved record within the app, particularly for committee member names. This is already done for author names using ORCID, and this data would integrate into the proposed app.
- Program the various forms throughout DAME interfaces to autofill possible entities that are being referred to
- If name is new to system, it will not be able to be handled automatically in version 1.0.
- A local data file for those names not found in LDAP, Scholars@TAMU or another external name sources will need to be created. This will be very important for legacy documents, such as the materials found in the Agricultural bulletins, leaflets, etc from the 20th century.
- Have something on the back-end to log/flag new names. Have a workflow to reconcile. This will directly impact the workflow for Metadata Management.
- Adding a new name to DAMName system when name added to LDAP or waiting until the person in LDAP has material with in the repository?
- Have backup plan on technical side to cover those faculty who do not have a UIN. Our understanding is that a new faculty member may be here for two weeks before they are added to LDAP.

VIII. Estimating time/effort needed to implement the solution, including, if applicable, time to (1) produce and (2) maintain a local authority file

It is hard to give an accurate estimate of the time and effort needed in order to have a working app. We do suggest it take place according to the following outline

1. Work needed before Name App can be deployed
 - Develop name app and an ID minting app
 - Build a name association helper to integrate with the DAMName App. It will read in the names from Dspace, LDAP, & Scholars@TAMU to read the names to look for possible matches for confirmation.
 - Generate a list of possible matches based on string similarity
 - The resources needed for development of the full app
 - 2+ developers
 - 4-6 Sprints
2. Metadata Clean-up (1&2 happening concurrently)
 - Human eyes for manual corrections of names
 - Work needed to clean up DSpace - Priorities (will be on-going)

- Current Faculty publications
 - Repository Committee members
 - 1 month
 - Dissertations and Thesis (Current)
 - Repository committee members
 - 2-3 months
 - Collections that could wait - Ag extension, Cushing?
 - Repository committee/ Metadata and cataloging
 - On-going. This will be expected to take several months to a year
 - Resources needed
 - Metadata and Cataloging unit
 - Respository Committee members
 - 1-3 months
- 3. Possible DSpace sprints?
 - Digital Initiatives 1-2 developers
 - Depends on what is needed. Too hard to estimate
- 4. Going live
 - From the time the App is approved to going live it is expected to take approximately 9 months to a year

IX. Next Steps

The committee believes that the initial research and investigation into: (a) available apps such as UNT Name's and UH Cedar has been concluded and they do not meet our needs and; (b) Existing name authority standards and coverage has been concluded. The results of this investigation leads us to recommend the development of an app that will work more robustly for our DAME system. It will pull from existing/external name authority resources, as well as from the internal ones of LDAP and VIVO.

Phase two (as we are calling it) of DAMNames would specify the resources needed to develop a name up, how it would be designed and how it would work. Although it will initially focus on personal names, we believe it can be expanded to include corporate names as well. However, the overwhelming need for authority control of personal names must be addressed first. It will including building a prototype to work out issues that will undoubtedly arise when the theory meets the reality of our repository.

Some issues and thoughts we have had concerning the next phase:

- a. Build in something to read in all the names in OAKTrust/Dspace, read in the names in LDAP to look for possible matches.
- b. Human eyes will be needed for manual corrections of names
- c. Approach the content in OAKtrust by prioritizing collections to be cleaned up. Possible collections to be handled first are
 - Current Faculty publications
 - Dissertations and Thesis (Current)
 - Need to include/focus on Committee members
 - Collections that could wait - Ag extension, Cushing?
- d. Possible DSpace sprints

- e. Investigate author indexing in DSpace
- f. Repository Committee gives specs to DI on what DC fields need to be indexed and how

We recognize that version 1.0 (AKA the prototype) will have limitations.

- App is not going to clean up DSpace (that will involve a separate tool)
- The process of populating the app will involve cleaning up DSpace
- Medium size program to help catalogers do the job
- Need to deal with every mode of ingest
- It will need to sit at the discovery layer

X. How will this solution affect legacy data/assets as well as newly-produced/added materials?

The issue of legacy data and assets is a very complicated and tangled. In the OAKTrust system, metadata has been imputed by multiple players - mass ingestion by developers and metadata librarians with metadata that may or may not have been created by librarians. There are also individual metadata items that have been created directly by users inputting their own work utilizing the Manikan program within OAKtrust. Within different collections and even within the same collections, understanding of what a contributor is, what a subject is, even what constitutes a title has varied. As a result, creating consistency with legacy data and assets will require a difficult, intensive and sustained effort on the part of Metadata Management.

The committee recommends that legacy data within OAKTrust (which, until recently, was the only IR database utilized) be prioritized in the order of what most significantly impacts our users. This is a decision for the Repository Committee to make. It also needs to be understood that this must be part of on-going database maintenance. It also is recommended that a database maintenance plan for the DAME be developed to ensure that when changes of programming or policy are made that affects the IR, that legacy data be considered.

The DAMNames phase two will delve into the issue in more detail. It is recommended that once a name and/or entity is discoverable within the IR, that a DAMName record be utilized for it. Basically, that every metadata record that has any field that falls under the prevue of authority control will have a record minted in the DAMName app. Within the entity records found in the DAMName, there will be note fields in which changes and dates of those changes can be noted. In this regard, the records within the DAMName app will utilize standard NACO (Name Authority Cooperative) cataloging standards. This will help to build robustness within the system as well as serve a useful function for user understanding of the material the entity (either personal or corporate).

Conclusions

The DAMNames committee has examined external name authority resources as well as three existing name authority apps, while looking at two - UNT's Names and UH'S Cedar - in depth. We have come to the conclusion that due to the specialized needs of the Texas A&M Libraries, as well as it's highly customized DAME, it would be best to develop an in-house app. Because of the complicated nature of

name authority control (or "Disambiguation") and the need to serve not just OAKTrust, but Scholars@TAMU and our newer institutional repository platforms (e.g., Avalon and Fedora), the app must be customizable so that it can 'talk' to multiple in-house and out-of-house programs.

The committee believes that starting with personal names is best, since the need for identifying individual researchers is immediate and critical. However, we believe that once personal names have been added to the app, organizational names can be added as well. Subject headings records would be somewhat different. The proposed app will be harvesting and 'fetching' information from external name resources (such as ISNI, ORCID, etc.). For subject records, the external resources, such as thesauri and other sources of controlled vocabularies, will be different. Resources such as LCSH authorities, Getty Vocabularies and others can be used as resources to mint 'subject' records. However we recommend treating this step last, since we believe that name authority control must take precedence at the beginning.

To this end, the Committee submits this report as an acknowledgement that we have reached a turning point with regards to the DAMNames project. We will continue to meet in order to assist in the creation of a DAMName app which, we believe, will not only provide name disambiguation, but will also be an important part of the Discovery interface for the DAME.

Submitted on Tuesday, November 20, 2018

by

Charity Stokes, Metadata Management (Chair)

Tatyana Chubara, Metadata Management

James Creel, Digital Initiatives

Jeannette Ho, Metadata Management

Dong Joon Lee, Scholarly Communication

David Lowe, Scholarly Communication