




# **Name Disambiguation for Digital Collections: Planning a Linked Data App for Authority Control at Texas A&M University Libraries**

**Jeannette Ho**  
**Cataloging/Metadata Librarian**  
**Texas A&M University Libraries**

**May 11, 2019**  
**LD4 Conference on Linked Data in Libraries**



# The Problem with names in TAMU's digital repository (OAKTrust) : Lack of disambiguation!

- Same or different person?

Smith, John

Smith, John

- Is “Wang, P.” the same person as any of the following:

Wang, Ping

Wang, Peng

Wang, Pingshu

Wang, PS?

Now showing ite

Authors Name

Wagner, Joshu

Wagner, Jr., Al

Wagner, K. [16]

Wagner, Kevin [17]

Wagner, Kevin L [1]

Wagner, Kevin Lee [1]

Wagner, Kimberly Ann [1]

Wagner, Logan [1]

Wagner, Luke A. [1]

Wagner, M. [13]

Wagner, Marcia Lynn [1]

Wagner, Matt [1]

Wagner, Matthew [2]

Wagner, Matthew Wayne [1]

Wagner, Nicole DeAnne [1]

Wagner, Norbert [4]

Wagner, Norbert G. [2]

Wagner, Norbert [1]

Wagner, Peter [1]

# Same or Different Person?: Lack of consistency leads to conflict!

Subjects

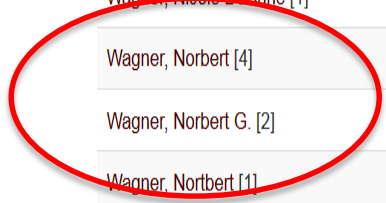
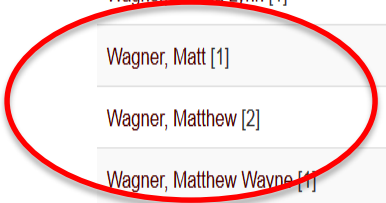

Department

MY ACCOUNT

Login

Register

Help and Documentation



# Design, Testing, And Field Experience Of A High-Pressure Natural Gas Reinjection Compressor.



View/Open  
t29pg039.pdf (1.902Mb)

Date  
2000  
Author  
Wagner, Norbert G.  
de Jongh, Frits M.  
Moffat, Robert

This paper explains the design and testing of a high-pressure centrifugal compressor. It also covers commissioning and field experience. Importance issues for high-pressure compressors are highlighted in the paper in general terms and for a specific design. Included are the mechanical impeller design, seal clearance control, and rotor stability. The impeller integrity is ensured by means of finite element analysis and supported by experimental modal analysis. During full load testing, a unique measurement was also carried out in order to verify the relative rotor displacement in the balance drum labyrinth under transient conditions. Special consideration is given to the verification of analytical rotor stability predictions, which is of major importance for high-pressure applications. The forces originating from the labyrinth seals tend to destabilize the rotor. The rotordynamic analysis is presented for a 418 bar (6061 psi) natural gas reinjection compressor. The predictions from this analysis were verified during an ASME/PTC-10 Class I (1997) test on the test stand. The compressor was tested with and without labyrinth swirl brakes installed resulted in reliable, trouble-free operation. This has been confirmed by actual field data from a floating production, storage, and offloading (FPSO) vessel in the North Sea.

**Clicking on this name will not link to any of his other "name variants" in OAKTrust:**  
-Wagner, Norbert  
-Wagner Nortbert [sic]

Metadata  
Show full item record

Turbomachines  
Collections  
Turbomachinery and Pump Symposia

Search [Q]

- Search OAKTrust
- This Collection

Advanced Search

BROWSE

- All of OAKTrust
- Communities & Collections
- By Issue Date
- Authors
- Titles
- Subjects


Subjects

Department

MY ACCOUNT

# The DAMEName working group (DAME=Digital Asset Management Ecosystem)

- **A working group was formed in 2018 to make recommendations for a “robust name authority system for Texas A&M affiliates and entities,” including:**
  - A standard approach (e.g., ISNI, ORCIDs), etc. alongside a newly minted URI-based identifier
  - A name application, if applicable
  - Basic technical needs for implementation
  - Estimated time/effort to implement the solution
  - Evaluate VIVO as a potential authority file system
  - Priorities among the need for authorities for A&M faculty members, staff, students, colleges, departments, etc.
- **While we did not rule out names of organizations, colleges and departments, etc. or subjects, our focus was on PERSONAL NAMES as the FIRST step**



# Activities of the working group (Phase I: June to December 2018, Phase 2: January 2019 to the present)

- **Conducted a literature review**
- **Reviewed existing standards for authority control and identifiers (focusing on persons)**
- **Examined existing “name apps”**
  - NAMES (University of North Texas)
  - CEDAR (University of Houston)
  - Others
- **Developed use cases (i.e., what would we like our app to do?)**
- **Recommended building our OWN app**
- **Explored and developed proposal for infrastructure and data sources for the app**

# Purposes of proposed app:

- **To serve as a tool for library personnel to manage identities in our DAME**
  - Every new name that gets input into IR should get minted with unique URI in the app (UUID)
  - Identify names for clean-up and reconciliation in the IR
- **To serve as a tool for metadata providers in OAKTrust and other repositories in our DAME to consistently select names that accurately identify and disambiguate authors**
- **To allow users of TAMU repositories to identify authors (and eventually organizational entities and subjects)**




# What should be stored in the app? At a MINIMUM, it should include:

- **Unique identifier (UUID)**
  - **“Canonical” name (“Murry, Robert D.)**
  - **Name variants (“Murray, Bob”, “Murray, Bobby”)**
  - **Links to external sources**
- We are actually managing IDENTITIES rather than NAMES!
- The unique URI (the UUID) is the central component that ties all name variants and links to external identifiers together




# Things we would like our “authority control” app to do:

- **Mint unique URI (using UUIDs) for each person in our repository**
  - That can be re-used by other components within our DAME (e.g., Fedora, Avalon, Spotlight, possibly FOLIO?, etc.).
  - Eventually, would like to link these URIs with ones associated with each digital “item” a person created or contributed to answer the question: “Show me all of this person’s works, despite form of name.”
- **Allow searching of both canonical name and name variants of each person or entity, in addition to retrieval via UUID**
  - App should cluster variants to answer the question: “Do I have all of this person’s works despite form of name”?
- **Disambiguate names for metadata providers and repository users so they can identify which name matches person being searched for**
  - To answer the question: “Is this the person that I am searching for?”
  - Use of contextual information



# Things we would like our “authority control” app to do (continued):

- **Identify names that need to be disambiguated (i.e., reconciled) after they are input into the IR**
- **Enable metadata providers to select appropriate names at the point of entering metadata to ensure consistency**
  - Via “type aheads” or drop-down lists that would provide choices that display the “canonical” form of name (plus additional contextual data?)
  - To do this, names in the app would need to be accessible to:
    - Authors who self-submit their works and metadata into OAKTrust (via Manakin self-submission form or Vireo ETD management software)
    - People who do NOT utilize a user interface when supplying metadata for the IR (e.g., spreadsheets for batch ingests)

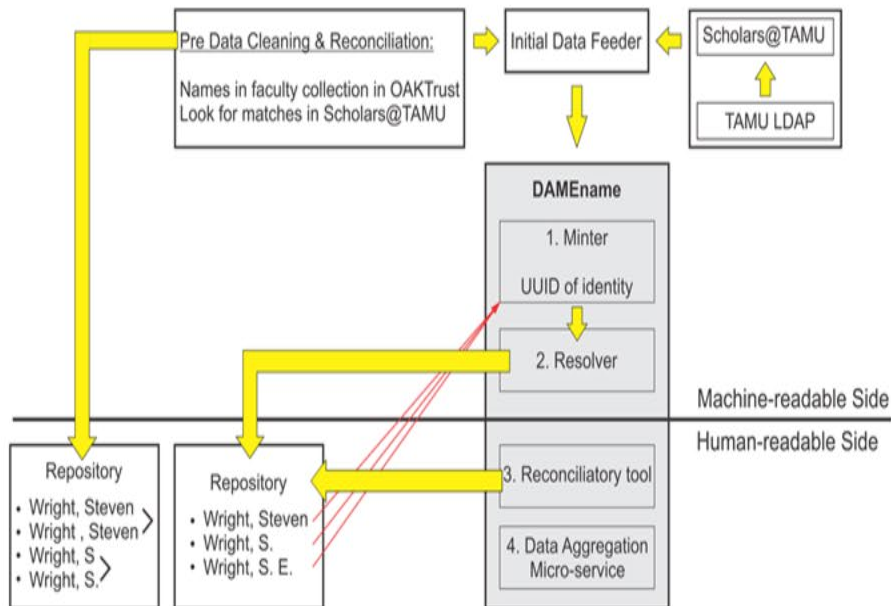


# Things we would like our “authority control” app to do (continued):

- **Our ultimate goal: Enable users (e.g., “the public”) to identify names of authors, etc. when searching our IR**
  - Via similar “type ahead” or drop-down list used for metadata providers
  - Need user interface that will allow them to access names in the app with enough disambiguating information for them to tell “This is the author I’m looking for” and retrieve all of their works.
    - The challenge of relying upon Solr indexing

## DAMName System Use Cases and Workflow

### The System Workflow: Initial Phase for Pilot



1. Minter: Assign a UUID for person
2. Resolver: Look up person by names or UUID and display aggregated contextual information about person
3. Reconciliation Process: Assign UUIDs to name variations to disambiguate identities
4. Data Aggregation Micro-Service: Pull contextual data to support the processes of resolution and reconciliation. This process is dependent on external

# “Data Aggregation Micro-service”: Where will disambiguating data come from?

- **It will link to multiple external sources of linked data, including Scholars (our local VIVO instance)**
- **Some possibilities that ranked the highest by the DAMEName working group:**
  - Library of Congress National Authority File (LCNAF)
  - ISNI (International Standard Identifier)
  - ORCID
  - VIAF (Virtual International Authority File)
  - Scopus
- **Other ones we looked at:**
  - Researcher ID
  - Wikidata
  - Google Scholar
  - MS Academic
  - Dimensions



# Criteria for including external identifiers in this service (Starred ones from Lee, D.J. & Stvilla, B. (2014))

- **Most important:**

- Uniqueness\*
- Persistence\*
- Compatibility\*
- Actionability/Resolvability\*

- **Other things we considered:**

- Interoperability\*
- Verifiability\*
- Controlling/gatekeeping body (e.g., PCC, ISNI Quality Team, etc.)
- API provided?
- Support RDF?
- Security\*
- Whether service is commercial vs. free
- Whether service is nonprofit

# SCHOLARS@TAMU

 Search

Herbert, Bruce | Professor

Co-author Network

### Positions

Map of Science

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

Co-investigator Network

### Contact Info

✉ [beherbert@tamu.edu](mailto:beherbert@tamu.edu)  
☎ 1 979 845 1083

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](#))

### Websites

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

### Research Areas

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

### ORCID

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

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


### education and training

- Ph.D. in Soil Science, [University of California, Riverside](#) 1992
- M.S. in Soil Science, [University of California, Riverside](#) 1988
- B.A. in Chemistry, [Colgate University](#) 1982

browse Any Phrase (APH:) sort by name

steven  [less](#)

approximate search



User id: ISNI | [log in](#) | [logout](#)

shortlist	<b>title data</b>	search history	1
<b>results</b> search [or] ISN:0000000464363572   1 hits			


<b>labels</b>	sources data	marc21
---------------	--------------	--------

**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

 **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





# Issues to be resolved in the future:

- **Should canonical name in the app be unique?**
  - Or should we just display names with contextual data pulled in from elsewhere?
- **What contextual data can be used for disambiguation?**
  - Some possibilities: date of birth, university affiliation data (college or department, title or status), subject areas, publication titles, etc.)
  - Will we control for colleges and departments? (establish as separate entities in the app?)
- **Where will contextual data used to disambiguate names come from?**
  - Some possibilities: VIVO Scholars database at TAMU, LDAP TAMU directory, external linked data sources (ORCID, ISNI, etc.).
  - Include “disambiguation” metadata field where info can be manually input (similar to University of North Texas’ name app)?



# Issues to be resolved (continued):

- **What will user interface look like for librarians/staff performing reconciliation? For metadata providers who will utilize the app?**
- **Where will canonical name come from for people not in TAMU's Scholars database or LDAP?**
  - LCNAF? Other external sources for identifiers? (ORCID, ISNI, etc.)?
  - What about people (e.g., historical, deceased persons) who aren't included in any of these sources? How can we do a "local" name authority for them?
- **How will app interact with the public interface of repositories within the TAMU DAME?**
  - App is meant as a tool for metadata providers and librarians who reconcile names, but can users of the IR benefit from contextual data it can "link out" to?
  - Maybe a link from name in IR to a knowledge card that utilizes links inside the app to pull in data from external sources?

# Issues to be resolved (continued):

- **How to deal with name changes?**
  - Treat as name variants or mint new identifier within the app for them?
  - Associate each name with dates when each name was used?
  - Persons vs. organizations



# Next steps:

- **Develop a prototype (projected date: July 2019)**
- **Test the prototype on selected collections**
  - Clean up names in collection prior to the test (typographical errors)
  - Initial test collections consist of names of TAMU faculty known to be in our Scholars/VIVO database:
    - Current faculty publications (estimated time: 1 month)
    - Current these and dissertations (estimated time: 2-3 months)
  - Run the names through the app
- **Analyze results of the test**
  - What are problems with the app that need to be tweaked?
  - What proportion of names need human review and reconciliation by metadata librarians/staff?
  - What are the frequencies of different types of problems? How much work does cleaning them up involve? (Implications for staffing & workflows)



# In the future, we hope to :

- **Test other collections**

- Including ones with names not in Scholars/VIVO database or LDAP, as well as older ones where names may not be represented in any external authorities
- How to handle historical names with no information to disambiguate them? (e.g., “Steward, T. S.”) Should we only disambiguate if no conflict?

- **Explore how information from app will look on the “human readable” side for metadata creators and IR users and how they will interact with it**
- **Plan how to deal with legacy data**
- **Eventually connect personal identities with their works in the IR**
- **Eventually include other types of entities in the app besides people (e.g., organizations, subjects)**

# Questions?



**My contact info: [jaho@library.tamu.edu](mailto:jaho@library.tamu.edu)**