ORGANIZATION NAME UNIFICATION PROCEDURES: UNIFYING ADDRESS VARIANTS

INTRODUCTION

A number of Clarivate discovery and benchmarking tools, such as InCites and WoS Organization Affiliation, draw upon data found in Web of Science publications. For effective use of such tools and accurate benchmarking, it is essential to correctly assign each WoS publication to an institution with which the author or authors is affiliated. This unification work is done by inspecting organization name variants found in publication addresses and building a record of known variants for each organization. The unification record then serves to:

- unify all current publications to the organization
- establish rules for attributing future indexed publications to the organization
- holds organization metadata such as country, region, address, alternate names, website URL, and more.

Unification of an organization's publication address variants is not completely automated. Rather, the process is completed through the use of proprietary software and the gathering and manual inspection of addresses. Clarivate unification team members utilize a number of tools and resources to understand the nature of an organization and how it is represented in WoS addresses:

- the full text of the WoS article itself
- the organization website
- authoritative organization resources such as Hoovers.com for corporations or the uniRank university ranking
- regional resources such as country ministries or other government sites
- Clarivate sales team members
- customer feedback

This document serves two purposes:

1. Explains the internal Clarivate process utilized to assign publications to the correct institution
2. Provides customers with guidance on how to submit address name variants for accurate and efficient creation of an organization's unification record

BENEFITS OF UNIFYING ORGANIZATION ADDRESS NAME VARIANTS

- Full accounting of all publications to ensure accurate data in InCites and related benchmarking tools
- Establishment of a WoS Organization Affiliation record that allows for easier discovery of publications in Web of Science
- Publication of accurate organization metadata in Clarivate products and services
- Development of an organization record that contains rules guiding how future publications are indexed

- Full transparency of the publication indexing and attribution process
- Accurate accounting of organization publications for use in various benchmarking activities

HOW UNIFICATION WORKS

In its most basic form, the overall unification process includes:

- Extraction of all records from the Web of Science for a particular geographic region, e.g. city or country, and extraction of all addresses from these papers.
- From each address, address segment data is extracted, normalized and ordered, including the main organization name, departmental names, the street address, and a geographical location.

In this way, this address:

Department of Psychology, Louisiana State University, Baton Rouge, LA 70803, United States

Is normalized becomes this address:

LOUISIANA STATE UNIV | DEPT PSYCHOL | BATON ROUGE | LA 70803 USA | USA

- Normalized address data is then ingested into Clarivate's unification system.
- From within the unification system, beginning with the most commonly occurring organization variants, the variants are searched and gathered into an organization unification record. The record unifies all current papers to the organization and establishes rules for how future indexed papers will be ingested and unified.

ADDRESS DATA TREATMENT

Organization affiliations indicated on publication addresses are neither homogeneous, nor consistent. Address unification aims to account for these challenges:

- Institution names within an address are not published in a consistent manner, making it difficult to accurately evaluate the data.
- Researchers utilize varying rules for designating the institution name; the rules of signature from one journal to another may differ.
- New organizations are created; organizations cease to exist in their current form; new organizational relationships are created; organizations merge or divest.
- An address may have been indexed only with ambiguous variants, e.g.,
  - Univ Technol
  - City Hosp
  - Naval Ctr
- A single address may refer to two or more separate institutions, e.g.,
  - AIM CEA DSM CNRS Univ Paris Diderot, Marseille, France
  - Univ Toronto York Univ Joint Program, 42 Saint George St, Toronto, Canada
- A single organization may have multiple affiliated organizations.

A number of practices and techniques are utilized to meet these address content disambiguation challenges:

- Other address data such as sub-organization name, department, acronym or postal code, is used to confirm the identity of the organization and correctly assign the address to all appropriate entities.
- Research is conducted on organizations referenced in addresses to understand the nature of the organization, its structure, its location and its affiliates.
- A network of internal and external advisors with specific knowledge of a geographic area or particular sector is consulted to discern the nature of an organization or group
- Customers participate in the unification process by analyzing addresses in their WoS publications and providing Clarivate with information about the addresses and the nature of the organization. Customer feedback comes in many forms:
  - the name and location of an organization
  - a list of known name variants without location context
  - a detailed list of known name variants, including a specific location context for each variant
  - a list of WoS accession numbers
  - a list if publication addresses

PREFERRED METHOD FOR RECEIVING CUSTOMER FEEDBACK

Customers may submit request for unification adjustments using the [Data Corrections form](#), and choosing "Organizations-enhanced" as the Change type. which provides space for your institution's basic contact information, as well as the ability to attach and upload a file. Data can be submitted as an attachment in most any format, including Excel, Word or similar.

Please submit as much publication data as possible, including

- organization name and location
- standard organization addresses (in all locations where institution operates)
- publication accession numbers, unique title numbers or DOI numbers
- standard organization name variants