

Web of Science Data Dictionary

Robert Light

IUNI/CNS

Version: January 22, 2016

General Database information

Database name: wos_core

This database is stored in the PostgreSQL instance on the Data Enclave

All data is the public schema

The Database Schema Map is available on the Enclave at /WoS/Documentation/Entity Relationship Diagram wos_core.pdf

From the terminal, the database can be accessed via psql (command line) or pgadmin3 (GUI). The Data Steward provides login information at account creation.

Total Records: 56,442,146

Example Structure

Table Name

Parent Table: Name of Parent Table

Total Records: Total number of expected records in the table

Defining XML Tag: The path of the XML tag that defines the table. For each instance of this tag in the record, a new row will be added to the table.

Table Structure:

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>table_id</i>	int4	Order of this entry in the record (internal primary key)	None
column1	varchar	The comment for this column in the database	(base tag value)
column2	varchar	These are usually taken from the XML schema documentation	@attribute_value
column3	varchar		child_tag_value
column4	varchar	Highlighted columns are empty	child/@attribute

Additional Notes:

Other known information about the table.

Column names have been chosen to reasonably match the tag and attribute names, though some changes have been made in the interest of brevity or to avoid confusion.

All raw data values are stored in the varchar or string format. XML enforces no type restrictions, so there is no guarantee that a column that is expected to contain numeric or date values will conform completely. Only internally generated counter IDs are integers.

All data has been parsed as is. There has been no effort to identify or correct errors in the data. Typographic errors or nonsensical values in the raw data have been transcribed as provided by Thomson Reuters.

Tables are sorted in alphabetical order.

wos_abstract_paragraphs

Abstract paragraphs

Parent Table: wos_abstracts

Total Records: 31,733,725

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/abstracts/abstract/abstract_text/p

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>abstract_id</i>	int4	Order of abstract record (internal primary key)	None
<i>paragraph_id</i>	int4	Order of abstract paragraph record (internal primary key)	None
paragraph_label	varchar	Abstract paragraph label	@label
paragraph_text	varchar	Abstract paragraph text	(Base tag value)

Additional Notes:

wos_abstracts

Abstracts

Parent Table: wos_summary

Total Records: 22,976,244

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/abstracts/abstract

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>abstract_id</i>	int4	Order of abstract record (internal primary key)	None
abstract_language	varchar	Language of abstract	@lang_id
abstract_type	varchar	Abstract type	@type
provider	varchar	Abstract provider	@provider
copyright_information	varchar	Copyright information	copyright_information
paragraph_count	varchar	Number of paragraphs in abstract	abstract_text/@count

Additional Notes:

wos_addresses

Full Record Metadata Address Data

Parent Table: wos_summary

Total Records: 65,158,241

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
addr_type	varchar	Address type	@type
addr_no	varchar	Address number. Used to link a specific author/publisher to a specific address.	address_spec/@addr_no
full_address	varchar	Full address	address_spec/full_addresses
full_address_lang_id	varchar	Language of full address data	address_spec/full_addresses/@lang_id
organization_count	varchar	Organizations count	address_spec/organizations/@count
suborganization_count	varchar	Suborganizations count	address_spec/suborganizations/@count
url_type	varchar	URL type	address_spec/url_spec/@url_type
url_date_info	varchar	Unstructured URL date information	address_spec/url_spec/@date_info
url_create_date	varchar	URL create date	address_spec/url_spec/@create_date
url_revised_date	varchar	Last date the URL was revised	address_spec/url_spec/@revised_date
url_cited_date	varchar	Date the URL was cited	address_spec/url_spec/@cited_date
url	varchar	URL	address_spec/url_spec/url
laboratory	varchar	Laboratory	address_spec/laboratory
laboratory_lang_id	varchar	Language of laboratory data	address_spec/laboratory/@lang_id
street	varchar	Street address	address_spec/street
street_lang_id	varchar	Language of street address	address_spec/street/@lang_id

city	varchar	City	address_spec/street
city_lang_id	varchar	Language of city data	address_spec/street/@lang_id
province	varchar	Province	address_spec/province
province_lang_id	varchar	Language of province data	address_spec/province/@lang_id
state	varchar	U.S. State or Canadian province	address_spec/state
state_lang_id	varchar	Language of state data	address_spec/state/@lang_id
country	varchar	Country	address_spec/country
country_lang_id	varchar	Language of country data	address_spec/country/@lang_id
post_num	varchar	Postal number	address_spec/post_num
post_num_lang_id	varchar	Language of postal number	address_spec/post_num/@lang_id
name_count	varchar	Count of associated names	names/@count

Additional Notes:

Post number seems to have been moved into the addresses_zip table

TO DO: remove wos_addresses:email_addr_count and doi_count, not in the config map, may need removal from addresses and reprint as well?

wos_address_names

Name Data from Addresses

Parent Table: wos_addresses

Total Records: 55,1154,084

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name
/names/name

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>name_id</i>	int4	Order of name record (internal primary key)	None
role	varchar	Role of the person or entity identified by the name	@role
seq_no	varchar	Sequence number of name in list	@seq_no
addr_no_raw	varchar	Raw address numbers, may be more than one separated by space	@addr_no
reprint	varchar	Reprint author flag	@reprint
lang_id	varchar	Language of the name	@lang_id
r_id	varchar	ResearcherID identifier	@r_id
r_id_tr	varchar	ResearcherID identifier as captured by TR	@r_id_tr
orcid_id	varchar	ORCID identifier	@orcid_id
orcid_id_tr	varchar	ORCID identifier as captured by TR	@orcid_id_tr
dais_id	varchar	Distinct Author Identification System identifier	@dais_id
display	varchar	Display flag. Used to indicate whether a name is displayed in full record.	@display
display_name	varchar	Name as displayed in record retrieved from database	display_name
full_name	varchar	Full name	full_name
wos_standard	varchar	Name in Web of Science standard format	wos_standard
prefix	varchar	Name prefix	prefix
first_name	varchar	First (given) name	first_name
middle_name	varchar	Middle name	middle_name
initials	varchar	Initials	initials
last_name	varchar	Last name (surname)	last_name
suffix	varchar	Name suffix	suffix

Additional Notes:

Roles breakdown:

role	count
author	55,100,486
corp	13,598

Reprint breakdown:

reprint	count
blank (presumed No)	44,049,191
Y	11,064,893

Coverage of identifiers

identifier	count (of non-NULL fields)
r_id	3,097,964
dais_id	15,754,187

wos_address_names_email_addr

Email Data from Address Names

Parent Table: wos_address_names

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name
/names/name/email_addr

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>name_id</i>	int4	Order of name record (internal primary key)	None
<i>email_id</i>	int4	Order on email record (internal primary key)	None
email_addr	varchar	Email address	(base tag value)
lang_id	varchar	Language ID	@lang_id

Additional Notes:

wos_address_organizations

Organization Data from Addresses

Parent Table: wos_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name
/organizations/organization

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>Id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>org_id</i>	int4	Order of organization record (internal primary key)	None
organization	varchar	Organization	(base tag value)
lang_id	varchar	Language ID of organization data	@lang_id

Additional Notes:

wos_address_suborganizations

Suborganization Data from Addresses

Parent Table: wos_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name
/suborganizations/suborganization

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>Id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>org_id</i>	int4	Order of suborganization record (internal primary key)	None
organization	varchar	Suborganization	(base tag value)
lang_id	varchar	Language ID of suborganization data	@lang_id

Additional Notes:

wos_address_zip

ZIP Data from Addresses

Parent Table: wos_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/addresses/address_name /zip

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>zip_id</i>	int4	Order of ZIP record (internal primary key)	None
zip	varchar	Zip code or postal code	(base tag value)
lang_id	varchar	Language ID	@lang_id
location	varchar	Undocumented	@location

Additional Notes:

wos_book_desc

Book description

Parent Table: wos_summary**Total Records:****Defining XML Tag:** /records/REC/static_data/item/book_desc

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>desc_id</i>	int4	Order of book description record (internal primary key)	None
bk_binding	varchar	Book binding	bk_binding
bk_publisher	varchar	Book publisher	bk_publisher
amount	varchar	Book price amount	bk_price/amount
currency	varchar	Currency of book price	bk_price/currency
price_desc	varchar	Book price description	bk_price/desc
price_volumes	varchar	Book price volumes - price of specified number of volumes	bk_price/volumes
bk_prepay	varchar	Book prepayment information	bk_prepay
bk_ordering	varchar	Book ordering information	bk_ordering

Additional Notes:

wos_book_desc

Book notes

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/item/book_notes

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>note_id</i>	int4	Order of book note record (internal primary key)	None
book_note	varchar	Book note	bk_binding

Additional Notes:

wos_conference

Summary Conference Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/summary/conferences/conference

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>conf_id</i>	varchar	Conference identifier	@conf_id
<i>conf_info_count</i>	varchar	conf_info count. There may be multiple variations of conference information for one conference.	conf_infos/@count
<i>conf_title_count</i>	varchar	Conference title count	conf_titles/@count
<i>conf_date_count</i>	varchar	Conference date count	conf_dates/@count
<i>conf_location_count</i>	varchar	Conference location count	conf_locations/@count
<i>sponsor</i>	varchar	Conference sponsor count	sponsors/@count
<i>conf_type</i>	varchar	Undocumented	conf_aux/@type
<i>lang_id</i>	varchar	Undocumented	conf_aux/@lang_id

Additional Notes:

wos_conf_date

Summary Conference Date Data

Parent Table: wos_conference

Total Records:

Defining XML Tag: /records/REC/static_data/summary/conferences/conference/conf_dates/conf_date

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>date_id</i>	int4	order of conf date record (internal primary key)	None
conf_date	varchar	Conference date	(base tag value)
conf_start	varchar	Conference start date	@conf_start
conf_end	varchar	Conference end date	@conf_end
display_date	varchar	Display date from source item	@display_date
lang_id	varchar	Language of conference date	@lang_id

Additional Notes:

wos_conf_info

Summary Conference Info Data

Parent Table: wos_conference

Total Records:

Defining XML Tag: /records/REC/static_data/summary/conferences/conference/conf_infos/conf_info

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>info_id</i>	int4	order of conf info record (internal primary key)	None
conf_info	varchar	Concatenation of conference title, conference date and conference location	(base tag value)
lang_id	varchar	Undocumented	@lang_id

Additional Notes:

wos_conf_location

Summary Conference Date Data

Parent Table: wos_conference

Total Records:

Defining XML Tag:

/records/REC/static_data/summary/conferences/conference/conf_locations/conf_location

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>location_id</i>	int4	order of conf location record (internal primary key)	None
composite_location	varchar	Full conference location	composite_location
composite_lang_id	varchar	Language of conference location	composite_location/@lang_id
conf_host	varchar	Conference host location	conf_host
host_lang_id	varchar	Language of conference host data	conf_host/@lang_id
conf_city	varchar	Conference city	conf_city
city_lang_id	varchar	Language of conference city data	conf_city/@lang_id
conf_state	varchar	U.S. State or Canadian province	conf_state
state_lang_id	varchar	Language of conference state data	conf_state/@lang_id

Additional Notes:

wos_conf_sponsor

Summary Conference Sponsor Data

Parent Table: wos_conference

Total Records:

Defining XML Tag: /records/REC/static_data/summary/conferences/conference/sponsors/sponsor

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>sponsor_id</i>	int4	order of conf sponsor record (internal primary key)	None
sponsor	varchar	Conference sponsor	(base tag value)
lang_id	varchar	Language of conference sponsor data	@lang_id

Additional Notes:

wos_conf_title

Summary Conference Title Data

Parent Table: wos_conference

Total Records:

Defining XML Tag: /records/REC/static_data/summary/conferences/conference/conf_titles/conf_title

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>conf_record_id</i>	int4	Order of conference record (internal primary key)	None
<i>title_id</i>	int4	order of conf title record (internal primary key)	None
conf_title	varchar	Conference title	(base tag value)
lang_id	varchar	Language of conference title	@lang_id

Additional Notes:

wos_contributors

Contributors

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/contributors/contributor

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>name_id</i>	int4	Order of name record (internal primary key)	None
role	varchar	Role of the person or entity identified by the name	@role
seq_no	varchar	Sequence number of name in list	@seq_no
addr_no_raw	varchar	Raw address numbers, may be more than one separated by space	@addr_no
reprint	varchar	Reprint author flag	@reprint
lang_id	varchar	Language of the name	@lang_id
r_id	varchar	ResearcherID identifier	@r_id
r_id_tr	varchar	ResearcherID identifier as captured by TR	@r_id_tr
orcid_id	varchar	ORCID identifier	@orcid_id
orcid_id_tr	varchar	ORCID identifier as captured by TR	@orcid_id_tr
dais_id	varchar	Distinct Author Identification System identifier	@dais_id
display	varchar	Display flag. Used to indicate whether a name is displayed in full record.	@display
display_name	varchar	Name as displayed in record retrieved from database	display_name
full_name	varchar	Full name	full_name
wos_standard	varchar	Name in Web of Science standard format	wos_standard
prefix	varchar	Name prefix	prefix
first_name	varchar	First (given) name	first_name
middle_name	varchar	Middle name	middle_name
initials	varchar	Initials	initials
last_name	varchar	Last name (surname)	last_name
suffix	varchar	Name suffix	suffix

Additional Notes:

wos_doctypes

Summary Document Type Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/summary/doctypes/doctype

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>doctype_id</i>	int4	Order of document type record (internal primary key)	None
doctype	varchar	Document type	(base tag value)
code	varchar	Undocumented	@code

Additional Notes:

wos_dynamic_identifiers

Dynamically-defined Identifier Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/dynamic_data/cluster_related/identifiers/identifier

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>dynamic_id</i>	int4	Order of identifier record (internal primary key)	None
identifier_type	varchar	Identifier type	@type
identifier_value	varchar	Identifier value	@value
self_ind	varchar	Undocumented	@self_ind

Additional Notes:

wos_edition

Web Of Science database editions

Parent Table: wos_summary

Total Records: 63,430,461

Defining XML Tag: /records/REC/static_data/summary/EWUID/edition

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>Id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>edition_ctr</i>	int4	Order of edition record (internal primary key)	None
edition	varchar	Associated Database Edition	@value

Additional Notes:

While the XML schema defines the relationship between this table and the summary table as 0..∞, in practice there is only one wos_page entry per record.

All records are linked to at least one edition.

The edition is a short code provided by Thomson Reuters. Summary statistics are below.

Acronym	Name	Record Count
WOS.SCI	Science Citation Index Expanded	42,263,961
WOS.SSCI	Social Sciences Citation Index	7,690,154
WOS.ISTP	Conference Proceedings Citation Index – Science	7,291,457
WOS.AHCI	Arts & Humanities Citation Index	4,281,088
WOS.ISSHP	Conference Proceedings Citation Index - Social Sciences	564,970
WOS.BHCI	Book Citation Index - Science	452,559
WOS.IC	Index Chemicus	359,872
WOS.BSCI	Book Citation Index - Social Sciences and Humanities	307,091
WOS.CCR	Current Chemical Reactions	219,309

Name source: <http://ipscience-help.thomsonreuters.com/wosWebServicesLite/dbEditionsOptionsGroup/databaseEditionsWos.html>

wos_grant_ids

Grant Identifiers

Parent Table: wos_grants

Total Records:

Defining XML Tag:

/records/REC/static_data/fullrecord_metadata/fund_ack/grants/grant/grant_ids/grant_id

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>grant_id</i>	int4	Order of grant record (internal primary key)	None
<i>id_id</i>	int4	Order of grant record (internal primary key)	None
grant_identifier	varchar	Grant ID	(Base tag value)

Additional Notes:

wos_grants

Grant Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/fund_ack/grants/grant

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>grant_id</i>	int4	Order of grant record (internal primary key)	None
grant_info	varchar	Unparsed grant information	(base tag value)
grant_info_language	varchar	Language of grant information	grant_info
grant_agency	varchar	Grant agency	grant_info/@lang_id
grant_agency_language	varchar	Language of grant agency	grant_agency
grant_agency_preferred	varchar	Preferred agency flag	grant_agency/@lang_id
alt_agency_count	varchar	Alternate agency count	alt_agencies/@count
grant_id_count	varchar	Grant ID count	grant_ids/@count
country	varchar	Grant country	country
acronym	varchar	Grant Acronym	acronym
investigator	varchar	Principal investigator	investigator

Additional Notes:

wos_headings

High Level Subject Headings

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/category_info/headings/heading

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>heading_id</i>	int4	Order of heading record (internal primary key)	None
heading	varchar	High-level subject heading	(base tag value)

Additional Notes:

wos_keywords

Keywords

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/keywords/keyword

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>keyword_id</i>	int4	Order of keyword record (internal primary key)	None
keyword	varchar	Keyword	(base tag value)
keyword_language	varchar	Language of keyword	@lang_id

Additional Notes:

wos_keywords_plus

Keywords Plus. Keywords generated from the titles of cited references

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/item/keywords_plus/keyword

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>keyword_id</i>	int4	Order of keyword record (internal primary key)	None
keyword_plus	varchar	Keyword Plus	(base tag value)
keyword_language	varchar	Language of keyword	@lang_id

Additional Notes:

wos_languages

Full Record Metadata Language Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/languages/language

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>language_id</i>	int4	Order of language record (internal primary key)	None
language	varchar	Normalized language	(base tag value)
language_type	varchar	Undocumented	@type
status	varchar	Undocumented	@status

Additional Notes:

wos_normalized_doctypes

Full Record Metadata Normalized Document Type Data

Parent Table: wos_summary

Total Records:

Defining XML Tag:

/records/REC/static_data/fullrecord_metadata/normalized_doctypes/normalized_doctype

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>doctype_id</i>	int4	Order of document type record (internal primary key)	None
doctype	varchar	Normalized document type	(base tag value)
code	varchar	Undocumented	@code

Additional Notes:

wos_normalized_languages

Full Record Metadata Normalized Language Data

Parent Table: wos_summary

Total Records:

Defining XML Tag:

/records/REC/static_data/fullrecord_metadata/normalized_languages/normalized_language

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>language_id</i>	int4	Order of language record (internal primary key)	None
language	varchar	Normalized language	(base tag value)
language_type	varchar	Undocumented	@type
status	varchar	Undocumented	@status

Additional Notes:

wos_page

Pagination Data

Parent Table: wos_summary

Total Records: 56,442,146

Defining XML Tag: /records/REC/static_data/summary/pub_info/page

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>page_id</i>	int4	Order of pagination record (internal primary key)	None
page_value	varchar	Pagination.raw_data text	(base tag value)
page_begin	varchar	Starting page number	@begin
page_end	varchar	Ending page number	@end
page_count	varchar	Page count	@count

Additional Notes:

While the XML schema defines the relationship between this table and the summary table as 0..∞, in practice there is only one wos_page entry per record.

The page_begin column is missing in 5.6% of entries and non-numeric in 2.7% of entries.

The page_end column is missing in 12.7% of entries and non-numeric in 2.7% of entries.

The page_count column is never missing and 100% numeric. It ranges from 0 to 90,003 with an average of 6.9. The page_count is zero in 6.4% of entries, though is uncertain whether these zeroes are intended to represent missing data, very short articles or articles of undefined length.

wos_publisher

Summary Publisher Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/summary/publishers/publisher

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>publisher_id</i>	int4	Order of publisher record (internal primary key)	None
addr_type	varchar	Address type	@type
addr_no	varchar	Address number. Used to link a specific author/publisher to a specific address.	address_spec/@addr_no
full_address	varchar	Full address	address_spec/full_addresses
full_address_lang_id	varchar	Language of full address data	address_spec/full_addresses/@lang_id
organization_count	varchar	Organizations count	address_spec/organizations/@count
suborganization_count	varchar	Suborganizations count	address_spec/suborganizations/@count
url_type	varchar	URL type	address_spec/url_spec/@url_type
url_date_info	varchar	Unstructured URL date information	address_spec/url_spec/@date_info
url_create_date	varchar	URL create date	address_spec/url_spec/@create_date
url_revised_date	varchar	Last date the URL was revised	address_spec/url_spec/@revised_date
url_cited_date	varchar	Date the URL was cited	address_spec/url_spec/@cited_date
url	varchar	URL	address_spec/url_spec/url
laboratory	varchar	Laboratory	address_spec/laboratory
laboratory_lang_id	varchar	Language of laboratory data	address_spec/laboratory/@lang_id
street	varchar	Street address	address_spec/street
street_lang_id	varchar	Language of street address	address_spec/street/@lang_id

city	varchar	City	address_spec/street
city_lang_id	varchar	Language of city data	address_spec/street/@lang_id
province	varchar	Province	address_spec/province
province_lang_id	varchar	Language of province data	address_spec/province/@lang_id
state	varchar	U.S. State or Canadian province	address_spec/state
state_lang_id	varchar	Language of state data	address_spec/state/@lang_id
country	varchar	Country	address_spec/country
country_lang_id	varchar	Language of country data	address_spec/country/@lang_id
post_num	varchar	Postal number	address_spec/post_num
post_num_lang_id	varchar	Language of postal number	address_spec/post_num/@lang_id
name_count	varchar	Count of publisher names	names/@count

Additional Notes:

TO DO: remove wos_publisher:email_addr_count and doi_count, not in the config map, may need removal from addresses and reprint as well?

wos_publisher_names

Name Data from Publishers

Parent Table: wos_publisher

Total Records:

Defining XML Tag: /records/REC/static_data/summary/publishers/publisher/names/name

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>publisher_id</i>	int4	Order of publisher record (internal primary key)	None
<i>name_id</i>	int4	Order of name record (internal primary key)	None
role	varchar	Role of the person or entity identified by the name	@role
seq_no	varchar	Sequence number of name in list	@seq_no
addr_no_raw	varchar	Raw address numbers, may be more than one separated by space	@addr_no
reprint	varchar	Reprint author flag	@reprint
lang_id	varchar	Language of the name	@lang_id
r_id	varchar	ResearcherID identifier	@r_id
r_id_tr	varchar	ResearcherID identifier as captured by TR	@r_id_tr
orcid_id	varchar	ORCID identifier	@orcid_id
orcid_id_tr	varchar	ORCID identifier as captured by TR	@orcid_id_tr
dais_id	varchar	Distinct Author Identification System identifier	@dais_id
display	varchar	Display flag. Used to indicate whether a name is displayed in full record.	@display
display_name	varchar	Name as displayed in record retrieved from database	display_name
full_name	varchar	Full name	full_name
wos_standard	varchar	Name in Web of Science standard format	wos_standard
prefix	varchar	Name prefix	prefix
first_name	varchar	First (given) name	first_name
middle_name	varchar	Middle name	middle_name
initials	varchar	Initials	initials
last_name	varchar	Last name (surname)	last_name
suffix	varchar	Name suffix	suffix

Additional Notes:

wos_references

Reference Details

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/references/reference

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>ref_ctr</i>	int4	Order of reference record (internal primary key)	None
<i>ref_id</i>	varchar	Unique WoS Identifier (within a collection)	uid
<i>cited_author</i>	varchar	Cited Author	citedAuthor
<i>assignee</i>	varchar	Assignee	assignee
<i>year</i>	varchar	Year	year
<i>page</i>	varchar	Page	page
<i>volume</i>	varchar	Volume	volume
<i>cited_title</i>	varchar	Cited Title - Cited Title is only available is the cited reference clusters with a source item, or if a "full Cited Reference" capture provided one	citedTitle
<i>cited_work</i>	varchar	Cited Work	citedWork
<i>doi</i>	varchar	Digital Object Identifier	doi
<i>art_no</i>	varchar	Undocumented	art_no
<i>patent_no</i>	varchar	Patent Number	patent_no

Additional Notes:

Note that the *cited_author* does seem to consistently list only one author, so should not be taken to be a full author list. Author lists for WoS documents can be access via the *ref_id*.

wos_reprint_addresses

Item Record Reprint Address Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
addr_type	varchar	Address type	@type
addr_no	varchar	Address number. Used to link a specific author/publisher to a specific address.	address_spec/@addr_no
full_address	varchar	Full address	address_spec/full_addresses
full_address_lang_id	varchar	Language of full address data	address_spec/full_addresses/@lang_id
organization_count	varchar	Organizations count	address_spec/organizations/@count
suborganization_count	varchar	Suborganizations count	address_spec/suborganizations/@count
url_type	varchar	URL type	address_spec/url_spec/@url_type
url_date_info	varchar	Unstructured URL date information	address_spec/url_spec/@date_info
url_create_date	varchar	URL create date	address_spec/url_spec/@create_date
url_revised_date	varchar	Last date the URL was revised	address_spec/url_spec/@revised_date
url_cited_date	varchar	Date the URL was cited	address_spec/url_spec/@cited_date
url	varchar	URL	address_spec/url_spec/url
laboratory	varchar	Laboratory	address_spec/laboratory
laboratory_lang_id	varchar	Language of laboratory data	address_spec/laboratory/@lang_id
street	varchar	Street address	address_spec/street
street_lang_id	varchar	Language of street address	address_spec/street/@lang_id

city	varchar	City	address_spec/street
city_lang_id	varchar	Language of city data	address_spec/street/@lang_id
province	varchar	Province	address_spec/province
province_lang_id	varchar	Language of province data	address_spec/province/@lang_id
state	varchar	U.S. State or Canadian province	address_spec/state
state_lang_id	varchar	Language of state data	address_spec/state/@lang_id
country	varchar	Country	address_spec/country
country_lang_id	varchar	Language of country data	address_spec/country/@lang_id
post_num	varchar	Postal number	address_spec/post_num
post_num_lang_id	varchar	Language of postal number	address_spec/post_num/@lang_id
name_count	varchar	Count of publisher names	names/@count

Additional Notes:

TO DO: remove wos_addresses:email_addr_count and doi_count, not in the config map, may need removal from addresses and reprint as well?

wos_reprint_address_names

Name Data from Reprint Addresses

Parent Table: wos_reprint_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact/names/name

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>name_id</i>	int4	Order of name record (internal primary key)	None
role	varchar	Role of the person or entity identified by the name	@role
seq_no	varchar	Sequence number of name in list	@seq_no
addr_no_raw	varchar	Raw address numbers, may be more than one separated by space	@addr_no
reprint	varchar	Reprint author flag	@reprint
lang_id	varchar	Language of the name	@lang_id
<i>r_id</i>	varchar	ResearcherID identifier	@r_id
<i>r_id_tr</i>	varchar	ResearcherID identifier as captured by TR	@r_id_tr
<i>orcid_id</i>	varchar	ORCID identifier	@orcid_id
<i>orcid_id_tr</i>	varchar	ORCID identifier as captured by TR	@orcid_id_tr
<i>dais_id</i>	varchar	Distinct Author Identification System identifier	@dais_id
display	varchar	Display flag. Used to indicate whether a name is displayed in full record.	@display
display_name	varchar	Name as displayed in record retrieved from database	display_name
full_name	varchar	Full name	full_name
wos_standard	varchar	Name in Web of Science standard format	wos_standard
prefix	varchar	Name prefix	prefix
first_name	varchar	First (given) name	first_name
middle_name	varchar	Middle name	middle_name
initials	varchar	Initials	initials
last_name	varchar	Last name (surname)	last_name
suffix	varchar	Name suffix	suffix

Additional Notes:

wos_reprint_address_names_email_addr

Email Data from Reprint Address Names

Parent Table: wos_reprint_address_names

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact/names/name/email_addr

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>name_id</i>	int4	Order of name record (internal primary key)	None
<i>email_id</i>	int4	Order on email record (internal primary key)	None
email_addr	varchar	Email address	(base tag value)
lang_id	varchar	Language ID	@lang_id

Additional Notes:

wos_reprint_address_organizations

Organization Data from Reprint Addresses

Parent Table: wos_reprint_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact/organizations/organization

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>org_id</i>	int4	Order of organization record (internal primary key)	None
organization	varchar	Organization	(base tag value)
lang_id	varchar	Language ID of organization data	@lang_id

Additional Notes:

wos_reprint_address_suborganizations

Suborganization Data from Reprint Addresses

Parent Table: wos_reprint_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact/suborganizations/suborganization

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>org_id</i>	int4	Order of suborganization record (internal primary key)	None
organization	varchar	Suborganization	(base tag value)
lang_id	varchar	Language ID of suborganization data	@lang_id

Additional Notes:

wos_reprint_address_zip

ZIP Data from Reprint Addresses

Parent Table: wos_reprint_addresses

Total Records:

Defining XML Tag: /records/REC/static_data/item/reprint_contact/zip

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>addr_id</i>	int4	Order of address record (internal primary key)	None
<i>zip_id</i>	int4	Order of ZIP record (internal primary key)	None
zip	varchar	Zip code or postal code	(base tag value)
lang_id	varchar	Language ID	@lang_id
location	varchar	Undocumented	@location

Additional Notes:

wos_reviewed_authors

Item Record Authors of Reviewed Work Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/item/reviewed_work/rw_authors/rw_author

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>author_id</i>	int4	Order of author record (internal primary key)	None
author	varchar	Reviewed work author	(base tag value)

Additional Notes:

wos_reviewed_languages

Item Record Language of Reviewed Work Data

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/item/reviewed_work/languages/language

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>language_id</i>	int4	Order of language record (internal primary key)	None
language	varchar	Language	(base tag value)

Additional Notes:

wos_subheadings

Subject category subheadings

Parent Table: wos_summary

Total Records:

Defining XML Tag:

/records/REC/static_data/fullrecord_metadata/category_info/subheadings/subheading

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>subheading_id</i>	int4	Order of subheading record (internal primary key)	None
subheading	varchar	Subject category subheading	(base tag value)

Additional Notes:

wos_subjects

Subject categories

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/fullrecord_metadata/category_info/subjects/subject

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>subject_id</i>	int4	Order of subject record (internal primary key)	None
subject	varchar	Subject category	(base tag value)
ascatype	varchar	Defines the two collection of subject categories used to classify journals in Web of Knowledge	@ascatype
code	varchar	Undocumented	@code
edition	varchar	Undocumented	@edition

Additional Notes:

wos_summary

Summary Record Data

Root table

Total Records: 56,442,146

Defining XML Tag: /records/REC/

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>Id</i>	varchar	WOS Record ID (primary key)	UID
file_number	int4	Internal record of which file this record is taken from, used for QC purposes	None
coll_id	varchar	Collection ID	static_data/summary/EW UID/WUID
pubyear	varchar	Publication Year	static_data/summary/pub_info/@pubyear
season	varchar	Publication Season	static_data/summary/pub_info/@season
pubmonth	varchar	Publication Month	static_data/summary/pub_info/@pubmonth
pubday	varchar	Publication Day	static_data/summary/pub_info/@pubday
coverdate	varchar	Concatenation of pubyear and pubmonth	static_data/summary/pub_info/@coverdate
edate	varchar	Undocumented	static_data/summary/pub_info/@edate
vol	varchar	Volume	static_data/summary/pub_info/@vol
issue	varchar	Issue	static_data/summary/pub_info/@issue
voliss	varchar	Volume + issue	static_data/summary/pub_info/@voliss
supplement	varchar	Journal Supplement	static_data/summary/pub_info/@supplement
special_issue	varchar	Journal Special Issue	static_data/summary/pub_info/@special_issue
part_no	varchar	Journal part number	static_data/summary/pub_info/@part_no
pubtype	varchar	Undocumented	static_data/summary/pub_info/@pubtype
medium	varchar	Undocumented	static_data/summary/pub_info/@medium

model	varchar	Medline's article @ Pubmodel	static_data/summary/pub_info/@model
indicator	varchar	Undocumented	static_data/summary/pub_info/@indicator
inpi	varchar	Indicator that source of chemical reaction data is INPI (Institut national de la propriete industriellee)'	static_data/summary/pub_info/@inpi
is_archive	varchar	Archive record	static_data/summary/pub_info/@is_archive
city	varchar	City of publication	static_data/summary/pub_info/@city
country	varchar	Country of publication	static_data/summary/pub_info/@country
has_abstract	varchar	Undocumented	static_data/summary/pub_info/@has_abstract
sortdate	varchar	Undocumented	static_data/summary/pub_info/@sortdate
title_count	varchar	Title count (Number of source title variations)	static_data/summary/titles/@count
name_count	varchar	Name count	static_data/summary/names/@count
doctype_count	varchar	Document type count	static_data/summary/doctypes/@count
conference_count	varchar	Conference count	static_data/summary/conferences/@count
language_count	varchar	Language count	static_data/fullrecord_metadata/languages/@count
normalized_language_count	varchar	Normalized language count	static_data/fullrecord_metadata/normalized_languages/@count
normalized_doctype_count	varchar	Normalized document type count	static_data/fullrecord_metadata/normalized_doctypes/@count
descriptive_ref_count	varchar	Descriptive reference count. Used for qualitative reference count. e.g. "Many refs".	static_data/fullrecord_metadata/descriptive_ref_count
reference_count	varchar	Reference count	static_data/fullrecord_metadata/refs/@count
address_count	varchar	Address count	static_data/fullrecord_metadata/addresses/@count
headings_count	varchar	Heading count	static_data/fullrecord_metadata/category_info/headings/@count
subheadings_count	varchar	Subheading count	static_data/fullrecord_metadata/category_info/subheadings/@count

subjects_count	varchar	Subject count	static_data/fullrecord_metadata/category_info/subjects/@count
fund_ack	varchar	Text of funding acknowledgement	static_data/fullrecord_metadata/fund_ack/fund_text
grants_count	varchar	Grant count	static_data/fullrecord_metadata/fund_ack/grants/@count
grants_complete	varchar	Grants complete flag	static_data/fullrecord_metadata/fund_ack/grants/@complete
keyword_count	varchar	Keyword count	static_data/fullrecord_metadata/keywords/@count
abstract_count	varchar	Abstract count	static_data/fullrecord_metadata/abstracts/@count
item_coll_id	varchar	Item record collection ID	static_data/item/@coll_id
item_ids	varchar	ISI Document Solution identifier	static_data/item/ids
item_ids_avail	varchar	IDS availability flag	static_data/item/ids/@avail
bib_id	varchar	Concatenation of volume, issue, page span, date	static_data/item/bib_id
bib_pagecount	varchar	Page count of source publication	static_data/item/bib_pagecount
bib_pagecount_type	varchar	Undocumented	static_data/item/bib_pagecount/@type
reviewed_language_count	varchar	Count of languages of reviewed work	static_data/item/reviewed_work/languages/@count
reviewed_author_count	varchar	Count of authors of reviewed work	static_data/item/reviewed_work/rw_authors/@count
reviewed_year	varchar	Reviewed work year	static_data/item/reviewed_work/rw_year
keywords_plus_count	varchar	Keywords plus count	static_data/item/keyword_plus/@count
book_chapters	varchar	Book chapter count	static_data/item/book_chapters
book_pages	varchar	Book pages	static_data/item/book_pages
book_notes_count	varchar	Count of book notes	static_data/item/book_notes/@count
chapterlist_count	varchar	Listed chapters	static_data/item/chapterlist/@count
contributor_count	varchar	Contributor count	static_data/contributors/@count

Additional Notes:

wos_summary_names

Name Data from Record Summary

Parent Table: wos_summary

Total Records:

Defining XML Tag: /records/REC/static_data/summary/names/name

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>name_id</i>	int4	Order of name record (internal primary key)	None
role	varchar	Role of the person or entity identified by the name	@role
seq_no	varchar	Sequence number of name in list	@seq_no
addr_no_raw	varchar	Raw address numbers, may be more than one separated by space	@addr_no
reprint	varchar	Reprint author flag	@reprint
lang_id	varchar	Language of the name	@lang_id
r_id	varchar	ResearcherID identifier	@r_id
r_id_tr	varchar	ResearcherID identifier as captured by TR	@r_id_tr
orcid_id	varchar	ORCID identifier	@orcid_id
orcid_id_tr	varchar	ORCID identifier as captured by TR	@orcid_id_tr
dais_id	varchar	Distinct Author Identification System identifier	@dais_id
display	varchar	Display flag. Used to indicate whether a name is displayed in full record.	@display
display_name	varchar	Name as displayed in record retrieved from database	display_name
full_name	varchar	Full name	full_name
wos_standard	varchar	Name in Web of Science standard format	wos_standard
prefix	varchar	Name prefix	prefix
first_name	varchar	First (given) name	first_name
middle_name	varchar	Middle name	middle_name
initials	varchar	Initials	initials
last_name	varchar	Last name (surname)	last_name
suffix	varchar	Name suffix	suffix

Additional Notes:

wos_summary_names_email_addr3

Name Data from Record Summary

Parent Table: wos_summary_names

Total Records:

Defining XML Tag: /records/REC/static_data/summary/names/name/email_addr

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>name_id</i>	int4	Order of name record (internal primary key)	None
<i>email_id</i>	int4	Order of email record (internal primary key)	None
email_addr	varchar	Email address	(base tag value)
lang_id	varchar	Language of email address	@lang_id

Additional Notes:

wos_titles

Title Data

Parent Table: wos_summary

Total Records: 327,757,500

Defining XML Tag: /records/REC/static_data/summary/titles/title

Column Name (Keys in Italics)	Data Type	Comment	Related XML tag/attribute
<i>id</i>	varchar	WOS Record ID (primary key)	/records/REC/UID
<i>page_id</i>	int4	Order of title record (internal primary key)	None
<i>title_type</i>	varchar	Title type	@type
<i>lang_id</i>	varchar	Language identifier	@lang_id
translated	varchar	Translated flag (title has been translated into English)	@translated
non_english	varchar	Non-English flag (original title not in English)	@non_english
title	varchar	Title	(base tag value)

Additional Notes:

Nearly all items have titles.

Most items have item and source titles. It looks like the abbreviations are mostly from the editions related to journal articles. Book series is interesting, because it is too large to be linked to the book editions.

Title Type counts

Title Type	Record Count
item	56,442,136
source	56,441,624
abbrev 11	54,128,604
source abbrev	54,128,266
abbrev 29	54,128,049
abbrev iso	48,195,025
series	2,914,571
book series	2,914,387
book subtitle	44,202
book	18,825
book seriessub	1,811