

# 2022 IRIS UMETRICS Data Dictionary and Data Diagram



Nicholls, N., Brown, C., & Owen-Smith, J. (2022). Data Dictionary for IRIS UMETRICS  
2022 Data Release. Institute for Research on Innovation and Science (IRIS).  
<https://doi.org/10.21987/6A4R-BG28>

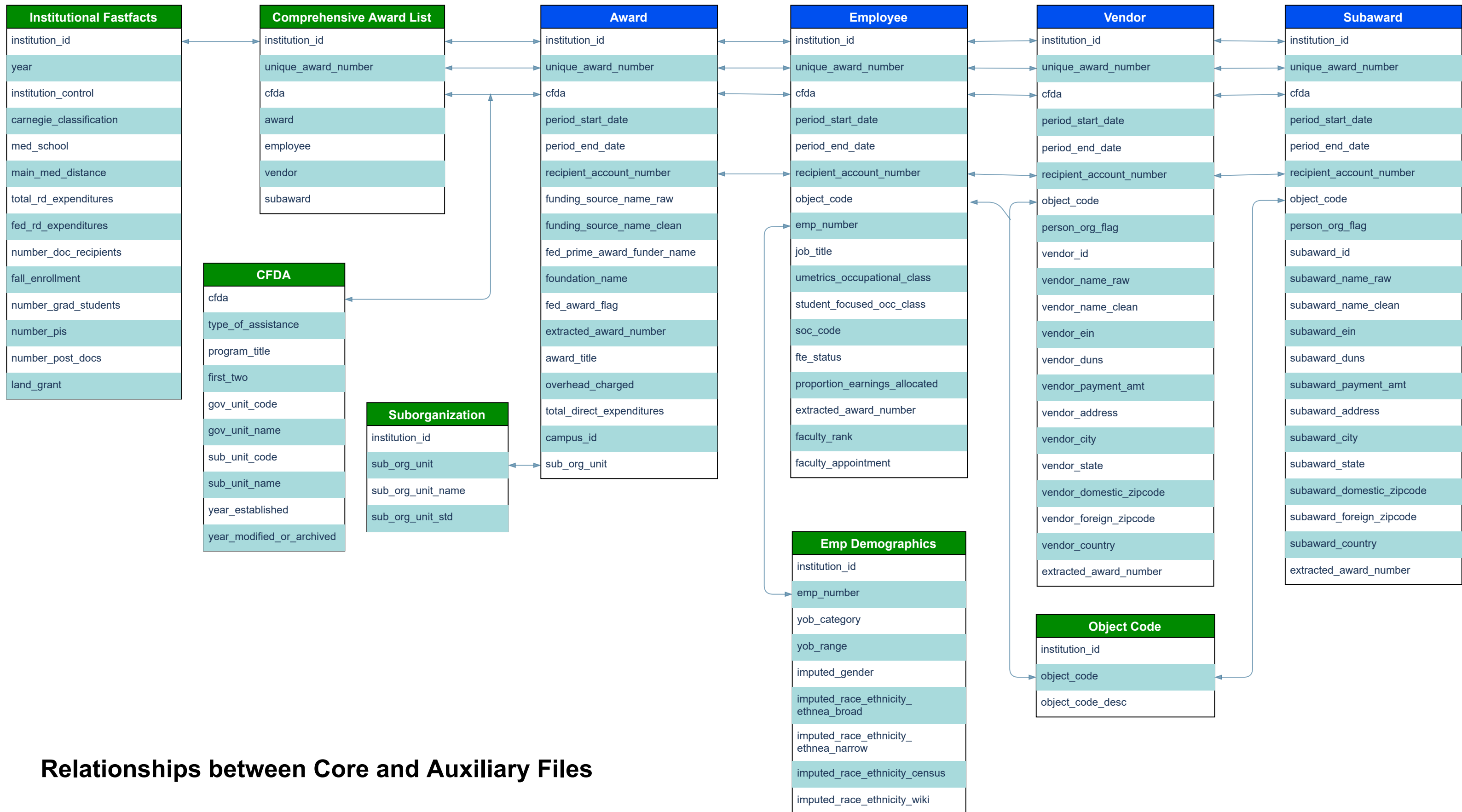
# Table of Contents

Dataset Structure .....	1
Data Diagram .....	2
Core Collection .....	9
Auxiliary Collection .....	13
Linkage Collection .....	16

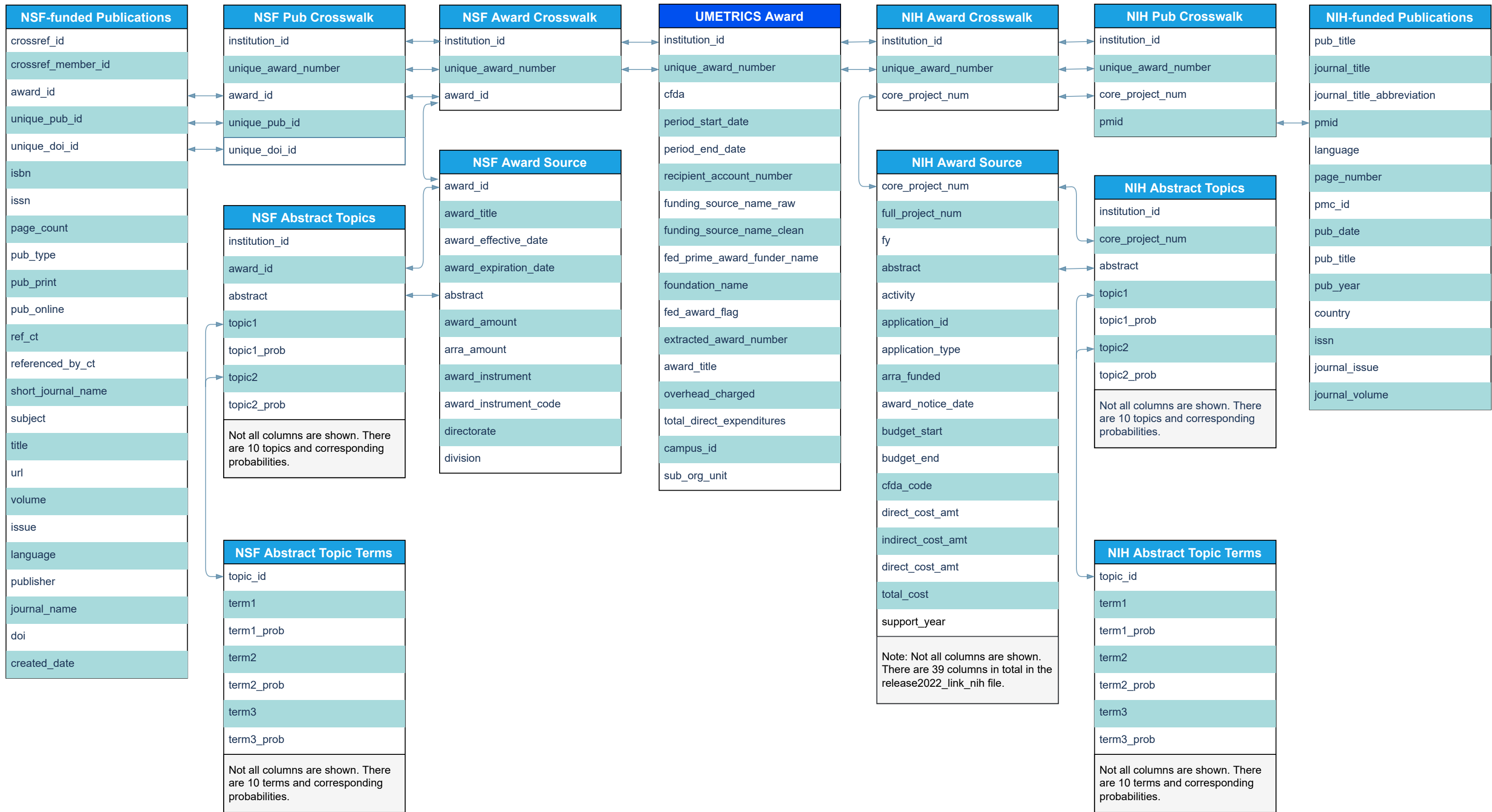
# IRIS UMETRICS 2022 Data Release

## Dataset Structure

Core Collection (4 Files)	Auxiliary Collection (6 Files)	Linkage Collection (21 Files)
file name	file name	file name
core_award	aux_cfda	link_doe_xwalk
core_employee	aux_comprehensive_award_list	link_dod_xwalk
core_vendor	aux_emp_demographics	link_nasa_xwalk
core_subaward	aux_institution_fastfacts	link_nih_xwalk
	aux_object_code	link_nsf_xwalk
	aux_suborg	link_usda_xwalk
		link_nih_pub_xwalk
		link_nsf_pub_xwalk
		link_pi_xwalk
		link_doe_source
		link_dod_source
		link_nasa_source
		link_nih_source
		link_nsf_source
		link_usda_source
		link_nih_pub_source
		link_nsf_pub_source
		link_nih_lda_abstract_topics
		link_nsf_lda_abstract_topics
		link_nih_lda_abstract_topic_terms
		link_nsf_lda_abstract_topic_terms



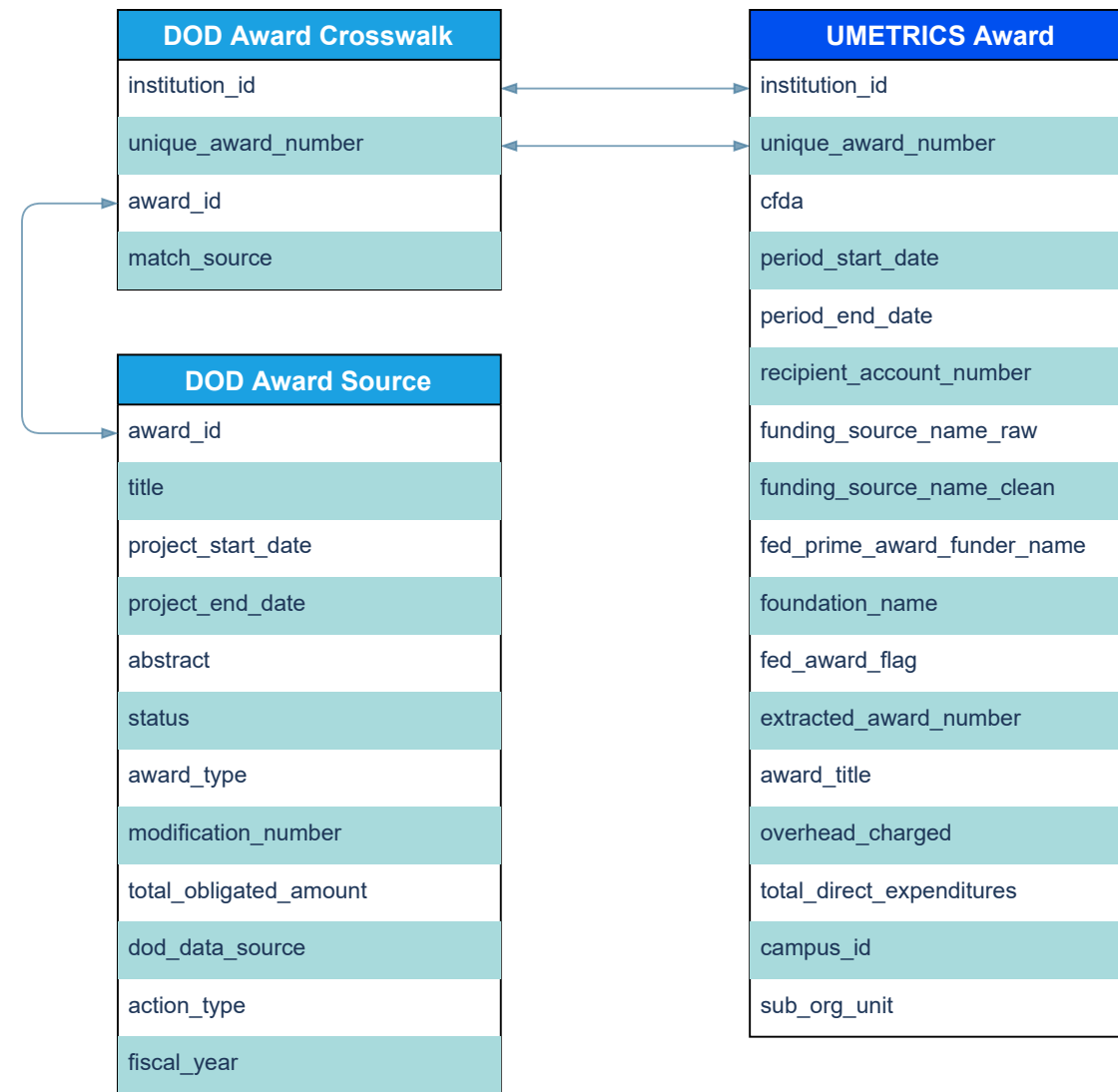
## Relationships between Core and Auxiliary Files



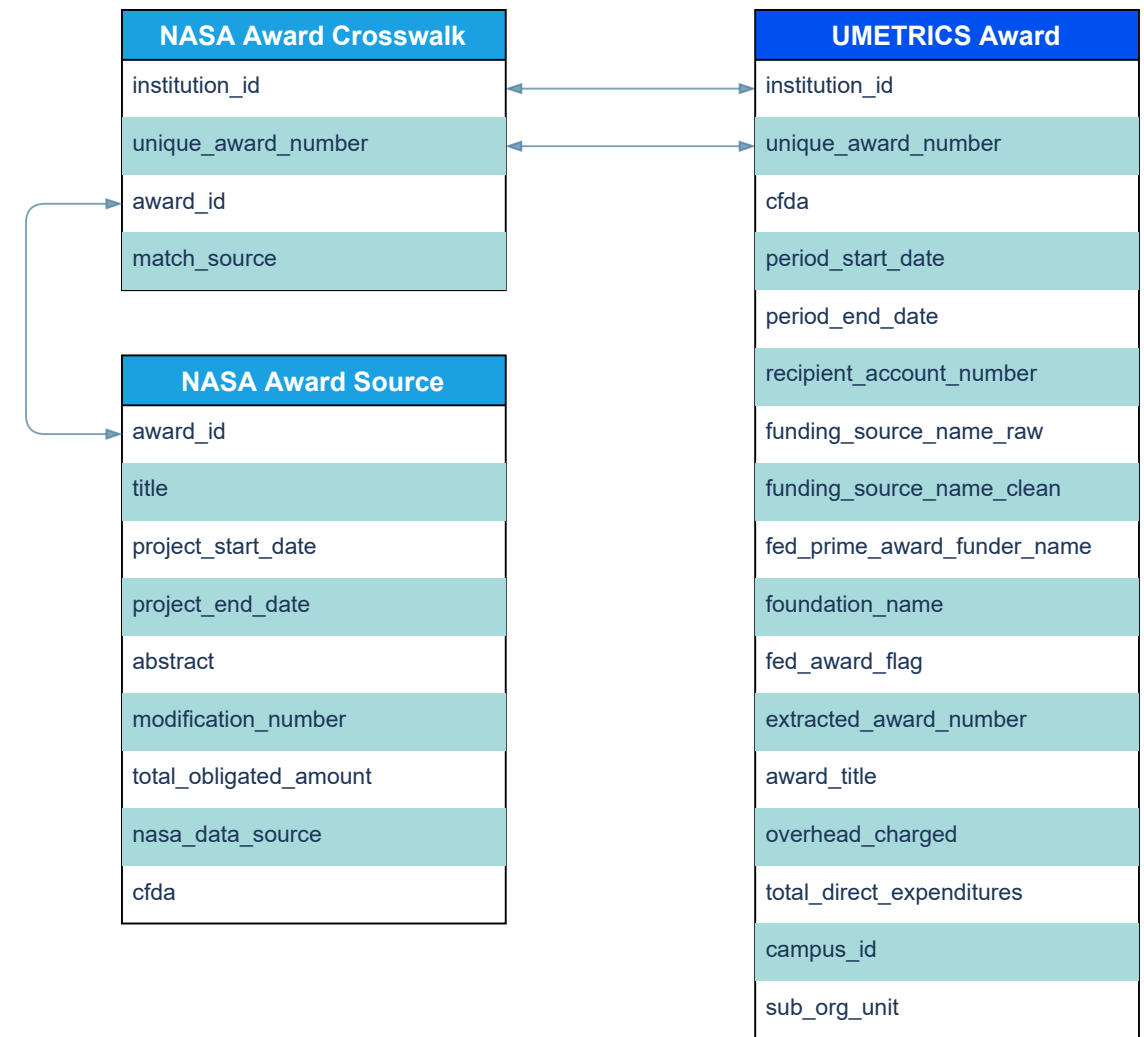
## Relationships between Core and Linkage Files (with a focus on NIH and NSF)



## Relationships between Core Award and Linkage Files (DOE)

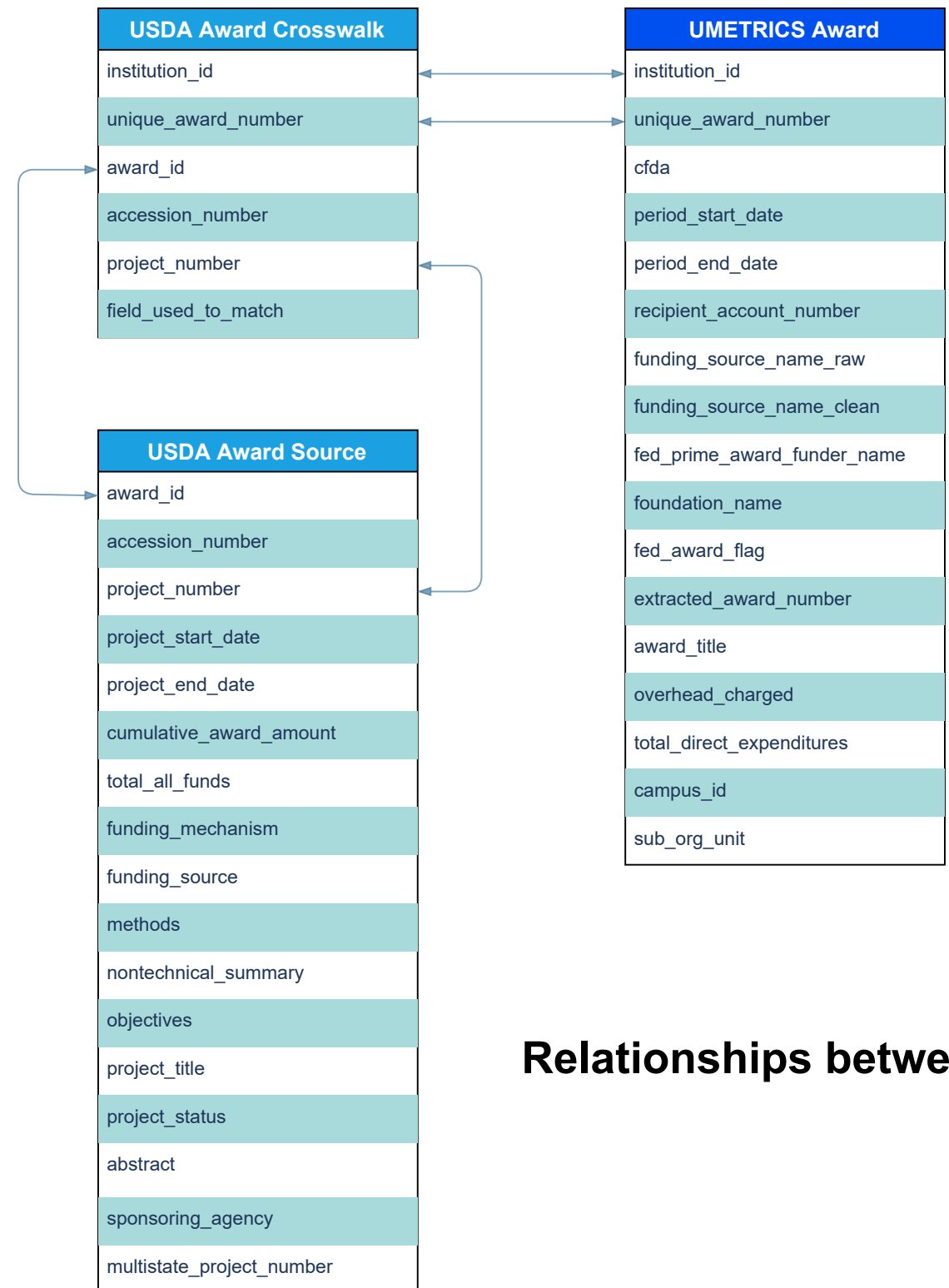


## Relationships between Core Award and Linkage Files (DOD)

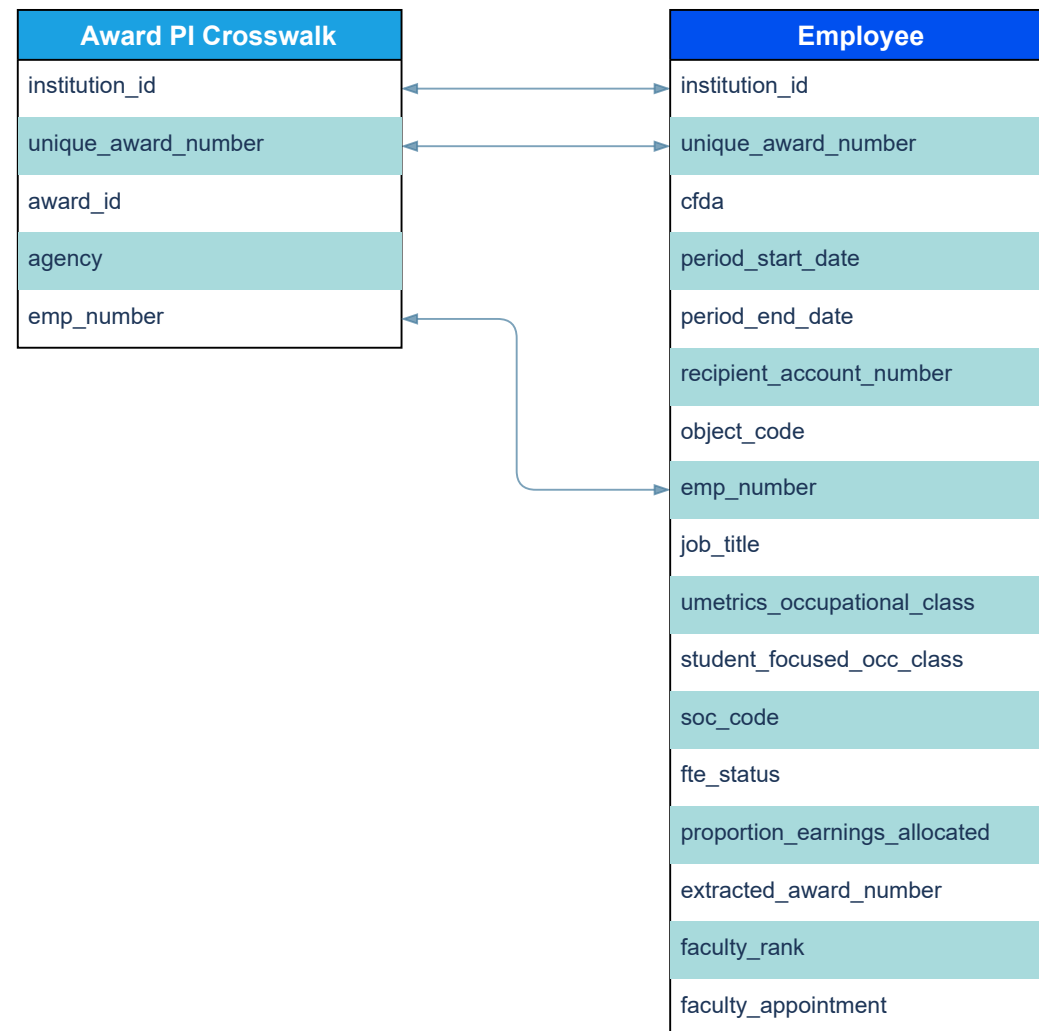


## Relationships between Core Award and Linkage Files (NASA)





## Relationships between Core Award and Linkage Files (USDA)



**Relationships between Core Employee and Linkage Files (Employees - Principal Investigators from six agencies)**

Files are organized by collection.  
Data fields are alphabetically ordered.

## core\_award

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Award Title	award_title	varchar	500	Title of award
2	Campus ID	campus_id	varchar	37	IRIS-generated identifier assigned to each campus for de-identification purposes. Each ID is a combination of institution_id and a serial number helpful to identify the campus to which a given award is made and by which the award is being administered
3	CFDA	cfda	varchar	10	A five-digit Catalog of Federal Domestic Assistance (CFDA) number retrieved from the unique_award_number that universities submit. This number is assigned to awards to represent the source of funding. If the first two digits are between 10 and 98, awards are federally funded. If the first two digits range between 00 and 09, or 99, sources of awards are most likely non-federal
4	Extracted Award Number	extracted_award_number	varchar	100	A substring of the 'unique_award_number' field; The CFDA number in the format of ##.### (or similar suffix part of the unique award number) is removed from the unique award number field, e.g., if the unique award number field shows "93.989 5-D31-AH9005-04", this extracted award number field shows "5-D31-AH9005-04"
5	Federal Award Flag	fed_award_flag	int	4	A binary code to identify types of awards based on the first two-digit number of CFDA or funding source name; coded 1 if the award is federal; coded 0 if non-federal; if the field federal prime award funder name field is not null, this field has the value of either 1 or 0, otherwise, null
6	Federal Prime Award Funder Name	fed_prime_award_funder_name	varchar	200	Name of federal sponsoring department or parent agency, e.g., Department of Health and Human Services for NIH awards; this is based on CFDA number or funding source name
7	Foundation Name	foundation_name	varchar	200	Disambiguated name of a funder that is a private agency or nonprofit / corporate foundation closely corresponding to the funding source name; this field is null if a funder is not identified as a private agency or foundation listed in the Foundation Directory Online by Candid (e.g., American Heart Association, Pfizer Foundation, American Cancer Society, Alfred P Sloan Foundation, Robert Wood Johnson Foundation, etc.)
8	Funding Source Name Clean	funding_source_name_clean	varchar	200	Cleaned name of funding source assigned to each project; if university is a recipient of a prime award, the funding source name is the direct sponsor; if university is a subrecipient of an award, this is the name of the pass-through entity (exceptions noted in documentation)
9	Funding Source Name Raw	funding_source_name_raw	varchar	200	Raw name of funding source assigned to each project; if university is a recipient of a prime award, the direct sponsor is the name of funding source; if university is a subrecipient of an award, this is the name of the pass-through entity (exceptions are noted in documentation)
10	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
11	Overhead Charged	overhead_charged	numeric	9	Overhead charged to the award in the specified period in actual dollars; Overhead charge is also commonly referred to as indirect cost
12	Period End Date	period_end_date	date	3	End of period in which a monthly expense transaction took place; each period end day is the last day of a month: e.g., 3/30/2008 or 12/31/2014
13	Period Start Date	period_start_date	date	3	Beginning of period in which a monthly expense transaction took place; each period start day is the first day of a month: e.g., 4/1/2009 or 10/1/2015
14	Recipient Account Number	recipient_account_number	varchar	50	A university's internal account number to uniquely identify each project; typically an accounting code used to allocate funds received from an award
15	Sub-organization Unit	sub_org_unit	varchar	62	IRIS-generated identifier assigned to sub-organization units to which each funded project is assigned, such as a particular college within a given IRIS member university. (This is not at the level of individual departments.) Each ID is a combination of campus_id described above and a serial number assigned to each sub-organization unit within each campus. This ID helps to identify the college or unit to which a given award is made and by which the award is being administered at a lower-level
16	Total Direct Expenditures	total_direct_expenditures	numeric	9	Total direct expenditures charged to the award in the specified period

17	Unique Award Number	unique_award_number	varchar	500	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
----	---------------------	---------------------	---------	-----	---

## core\_employee

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	CFDA	cfda	varchar	10	A five-digit Catalog of Federal Domestic Assistance (CFDA) number retrieved from the unique_award_number that universities submit. This number is assigned to awards to represent the source of funding. If the first two digits are between 10 and 98, awards are federally funded. If the first two digits range between 00 and 09, or 99, sources of awards are most likely non-federal
2	Employee Number	emp_number	varchar	200	IRIS-generated unique identifier assigned to all personnel paid by awards
3	Extracted Award Number	extracted_award_number	varchar	100	A substring of the 'unique_award_number' field; The CFDA number in the format of ###.### (or similar suffix part of the unique award number) is removed from the unique award number field, e.g., if the unique award number field shows "93.989 5-D31-AH9005-04", this extracted award number field shows "5-D31-AH9005-04"
4	Faculty Rank	faculty_rank	varchar	50	Faculty ranking has four categories: "Full Prof", "Associate Prof", "Assistant Prof", and "Unknown"; those who are identified as "Faculty" in the "UMETRICS occupational class" field are grouped into one of these categories
5	Faculty Appointment	faculty_appointment	varchar	50	Faculty appointment consists of four categories corresponding to the type of appointment: "tenure-track", "nontenure-track", "emeritus-retired", and "appointment unknown"; those who are identified as "Faculty" in the field of "UMETRICS occupational class" are grouped into one of these categories
6	FTE Status	fte_status	numeric	9	Designation of the status of funded personnel (e.g., full time = 1.0, half time = .5); FTE is a university specific, not an award specific field; the value ranges between 0 and 1
7	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
8	Job Title	job_title	varchar	200	Job or occupation title assigned to the funded personnel by IRIS member universities
9	Object Code	object_code	varchar	50	Internal object code or other expense type category assigned to a transaction to identify payment purposes or resources
10	Period End Date	period_end_date	date	3	End of period in which a monthly expense transaction took place; each period end day is the last day of a month: e.g., 3/30/2008 or 12/31/2014
11	Period Start Date	period_start_date	date	3	Beginning of period in which a monthly expense transaction took place; each period start day is the first day of a month: e.g., 4/1/2009 or 10/1/2015
12	Proportion of Earnings	proportion_earnings_allocated	numeric	9	Calculated portion of earnings charged by funded personnel to the award in the specified period. This is not actual salary or dollar amounts; the value ranges between 0 and 1 depending on how much of the salary is derived from an award
13	Recipient Account Number	recipient_account_number	varchar	50	A university's internal account number to uniquely identify each project; typically an accounting code used to allocate funds received from an award
14	SOC Code	soc_code	varchar	50	Standard Occupational Classification codes that are required for federal agency reporting ( <a href="http://www.bls.gov/soc/">http://www.bls.gov/soc/</a> ); each occupation in the SOC is placed within one of 23 major groups
15	Student-focused Occupational Class	student_focused_occupational_class	varchar	50	Student-focused Occupational Class is one of the two occupational classifications generated by IRIS; focusing on students, this field includes six groups: Undergraduate, Graduate Student, Other Student, Post Graduate Researcher, Faculty, and Staff; this classification is a result of analyzing a combination of university supplied job titles, occupational classes, CFDA numbers, and object code descriptions in order to provide a more accurate picture of student involvement in sponsored research. See documentation for more details on the definition of each category.

16	UMETRICS Occupational Class	umetrics_occupational_class	varchar	50	UMETRICS Occupational Class is one of the two occupational classifications generated by IRIS; focusing on faculty, jobs are categorized into 6 major aggregate groups (Faculty, Staff, Post Graduate Research, Graduate Student, Undergraduate, and Other). The Staff group is further classified into 6 categories (Clinical, Research, Research Facilitation, Technical Support, Instructional, Other Staff). See documentation for more details on the definition of each category.
17	Unique Award Number	unique_award_number	varchar	500	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)

## core\_vendor

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	CFDA	cfda	varchar	10	A five-digit Catalog of Federal Domestic Assistance (CFDA) number retrieved from the unique_award_number that universities submit. This number is assigned to awards to represent the source of funding. If the first two digits are between 10 and 98, awards are federally funded. If the first two digits range between 00 and 09, or 99, sources of awards are most likely non-federal
2	Extracted Award Number	extracted_award_number	varchar	100	A substring of the 'unique_award_number' field; The CFDA number in the format of ###.### (or similar suffix part of the unique award number) is removed from the unique award number field, e.g., if the unique award number field shows "93.989 5-D31-AH9005-04", this extracted award number field shows "5-D31-AH9005-04"
3	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
4	Object Code	object_code	cfda	50	Internal object code or other expense type category assigned to a transaction to identify payment purposes or resources
5	Period End Date	period_end_date	date	3	End of period in which a monthly expense transaction took place; each period end day is the last day of a month: e.g., 3/30/2008 or 12/31/2014
6	Period Start Date	period_start_date	date	3	Beginning of period in which a monthly expense transaction took place; each period start day is the first day of a month: e.g., 4/1/2009 or 10/1/2015
7	Person Organization Flag	person_org_flag	varchar	1	A binary code ('P' for person or 'O' for organization) to differentiate type of vendors. This dichotomous category was utilized to mask vendors' personally identifiable information
8	Recipient Account Number	recipient_account_number	varchar	50	A university's internal account number to uniquely identify each project; typically an accounting code used to allocate funds received from an award
9	Unique Award Number	unique_award_number	varchar	500	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
10	Vendor Address	vendor_address	varchar	200	Address of the vendor. IRIS has replaced vendor address with the string of 'masked' if vendor addresses are provided by universities (thus not null) and vendors are individuals
11	Vendor City	vendor_city	varchar	50	City of the vendor associated with the vendor address
12	Vendor Country	vendor_country	varchar	50	Country of the vendor associated with the vendor address
13	Vendor Domestic Zipcode	vendor_domestic_zipcode	varchar	50	US ZIP code of vendor associated with the vendor address
14	Vendor DUNS	vendor_duns	varchar	50	A vendor's nine-digit Data Universal Numbering System (DUNS) number to identify business entities on a location-specific basis is copyrighted and provided by Dun & Bradstreet (D&B). If a DUNS number was provided (i.e., not null) by universities for vendors that are individuals, IRIS has replaced DUNS number with hashed DUNS number for de-identification purposes
15	Vendor EIN	vendor_ein	varchar	50	A vendor's nine-digit Employer Identification Number (EIN). If an EIN was provided (i.e., not null) by universities for vendors that are individuals, IRIS has replaced EIN with hashed EIN for de-identification purposes

16	Vendor Foreign Zip Code	vendor_foreign_zipcode	varchar	50	Foreign ZIP/postal code of vendor associated with the vendor address
17	Vendor ID	vendor_id	varchar	200	IRIS-generated unique identifier assigned to the vendor (an organization or individual) that provides goods or services paid by an IRIS member university's award. IRIS cleans vendor name records from the data submitted by universities and generates this identifier based on the cleaned names
18	Vendor Name	vendor_name	varchar	500	Cleaned name of the vendor. IRIS has replaced vendor names with the string of 'masked' if names are provided by universities (thus not null) and vendors are individuals
19	Vendor Name Raw	vendor_name_raw	varchar	500	Raw name of the vendor. IRIS has replaced vendor names with the string of 'masked' if names are provided by universities (thus not null) and vendors are individuals
20	Vendor Payment Amount	vendor_payment_amt	numeric	9	Funds charged to the award by the vendor in the specified period; vendor payment amounts can be negative
21	Vendor State	vendor_state	varchar	50	State of the vendor associated with the vendor address

## core\_subaward

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	CFDA	cfda	varchar	10	A five-digit Catalog of Federal Domestic Assistance (CFDA) number retrieved from the unique_award_number that universities submit. This number is assigned to awards to represent the source of funding. If the first two digits are between 10 and 98, awards are federally funded. If the first two digits range between 00 and 09, or 99, sources of awards are most likely non-federal
2	Extracted Award Number	extracted_award_number	varchar	100	A substring of the 'unique_award_number' field; The CFDA number in the format of ##.### (or similar suffix part of the unique award number) is removed from the unique award number field, e.g., if the unique award number field shows "93.989 5-D31-AH9005-04", this extracted award number field shows "5-D31-AH9005-04"
3	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
4	Object Code	object_code	varchar	50	Internal object code or other expense type category assigned to a transaction to identify payment purposes or resources
5	Period End Date	period_end_date	date	3	End of period in which a monthly expense transaction took place; each period end day is the last day of a month: e.g., 3/30/2008 or 12/31/2014
6	Period Start Date	period_start_date	date	3	Beginning of period in which a monthly expense transaction took place; each period start day is the first day of a month: e.g., 4/1/2009 or 10/1/2015
7	Person Organization Flag	person_org_flag	varchar	1	A binary code ('P' for person or 'O' for organization) to differentiate type of subawardees. This dichotomous category was utilized to mask vendors' personally identifiable information
8	Recipient Account Number	recipient_account_number	varchar	50	A university's internal account number to uniquely identify each project; typically an accounting code used to allocate funds received from an award
9	Subaward Address	subaward_address	varchar	200	Address of the subawardee. IRIS has replaced subawardee address with the string of 'masked' if subawardee address are provided by universities (thus not null) and subawardees are individuals
10	Subaward City	subaward_city	varchar	50	City of the subawardee associated with the subawardee address
11	Subaward Country	subaward_country	varchar	50	Country of the subawardee associated with the subawardee address
12	Subaward Domestic Zipcode	subaward_domestic_zipcode	varchar	50	US ZIP code of subawardee associated with the subawardee address
13	Subaward DUNS	subaward_duns	varchar	50	A subawardee's nine-digit Data Universal Numbering System (DUNS) number to identify business entities on a location-specific basis is copyrighted and provided by Dun & Bradstreet (D&B). If a DUNS number was provided (i.e., not null) by universities for subawardees that are individuals, IRIS has replaced DUNS number with hashed DUNS number for de-identification purposes
14	Subaward EIN	subaward_ein	varchar	50	A subawardee's nine-digit Employer Identification Number (EIN). If an EIN was provided (i.e., not null) by universities for subawardees that are individuals, IRIS has replaced EIN with the string 'masked'
15	Subaward Foreign Zip Code	subaward_foreign_zipcode	varchar	50	Foreign ZIP/postal code of subawardee associated with the subawardee address
16	Subaward ID	subaward_id	varchar	200	IRIS-generated unique identifier assigned to the subaward recipient organization to which an IRIS member university provides program awards / subgrants / subcontracts. IRIS cleans subawardee name records from the data submitted by universities and generates this identifier based on the cleaned names

17	Subaward Name Clean	subaward_name_clean	varchar	500	Cleaned name of the subawardee. IRIS has replaced subawardee names with the string of 'masked' if names are provided by universities (thus not null) and subawardees are individuals
18	Subaward Name Raw	subaward_name_raw	varchar	500	Raw name of the subawardee. IRIS has replaced subawardee names with the string of 'masked' if names are provided by universities (thus not null) and subawardees are individuals
19	Subaward Payment Amount	subaward_payment_amt	numeric	9	Funds charged to the award by the subawardee in the specified period; subaward payment amounts can be negative
20	Subaward State	subaward_state	varchar	50	State of the subawardee associated with the subawardee address
21	Unique Award Number	unique_award_number	varchar	500	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)

## aux\_cfda

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	CFDA	cfda	varchar	10	A Catalog of Federal Domestic Assistance (CFDA) number used to identify and sort federal financial assistance programs. Each CFDA number contains five digits and appears in the following format: ##.### (e.g., 10.001 or 93.301)
2	First Two Digits of CFDA	first_two	int	4	The first two digits of CFDA numbers that appear in the CFDA field described above. The first two digit number can be used to map each federal financial assistance program to funding agencies. The first two digit numbers (in integer format) range between 10 and 98
3	Government Unit Code	gov_unit_code	varchar	50	The official acronym of the assisting federal agency at the highest level of a given domestic assistance program. For example, if a domestic assistance program is administered and managed by the National Institutes of Health, this field indicates its parent organization acronym, HHS
4	Government Unit Name	gov_unit_name	varchar	200	The name of the assisting federal agency at the highest level of a given domestic assistance program. For example, if a domestic assistance program is administered and managed by the National Institutes of Health, this field indicates its parent organization, the Department of Health and Health Sciences
5	Program Title	program_title	varchar	500	The program title that is available and downloaded from the website, cfda.gov (sam.gov). This field is null for CFDA's that are not found in the historical file. See above for the definition of historical file
6	Sub-unit Code	sub_unit_code	varchar	50	The official acronym of the federal agency that directly provides and administers a given domestic assistance program. For example, if a domestic assistance program is provided and administered by the National Institutes of Health, this field indicates NIH
7	Sub-unit Name	sub_unit_name	varchar	200	The name of the federal agency that directly provides and administers a given domestic assistance program. For example, if a domestic assistance program is provided and administered by the National Institutes of Health, this field indicates National Institutes of Health. NOTE: In the historical file, the equivalent field is named "office"
8	Type of Assistance	type_of_assistance	varchar	500	Each program is identified in terms of one or more of the 15 types of assistance provided. This field has values only for active CFDA's
9	Program Established Year	year_established	int	4	The year when a given domestic assistance program (associated with a unique CFDA number) was established. The field is null if CFDA numbers are not found in the historical file. See above for the definition of historical file
10	Program Modified or Archived Year	year_modified_or_archived	int	4	The year when a given domestic assistance program was modified or archived

## aux\_comprehensive\_award\_list

	Field Name	Column Name	Data Type	Set Length	Field Definition
--	------------	-------------	-----------	------------	------------------

1	Award File	award	int	4	A binary code to differentiate the file from which a given award originates; coded 1 if the award is present in Award file; coded 0 otherwise
2	CFDA	cfda	varchar	10	A five-digit Catalog of Federal Domestic Assistance (CFDA) number retrieved from the unique_award_number that universities submit. This number is assigned to awards to represent the source of funding. If the first two digits are between 10 and 98, awards are federally funded. If the first two digits range between 00 and 09, or 99, sources of awards are most likely non-federal
3	Employee File	employee	int	4	A binary code to differentiate the file from which a given award originates; coded 1 if the award is present in Employee file; coded 0 otherwise
4	Federal award flag	fed_award_flag	int	4	A binary code to identify federally funded awards; coded 1 if award is federal, 0 otherwise. This flag for each unique award number corresponds to the value in the 'fed_award_flag' field in the Award file if the unique award number appears in the Award file, in other words 0 in the Award field in this file; if a given unique award number does not appear in the Award file (thus 0 in the Award field), then this indicator is based on CFDA
5	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
6	Subaward File	subaward	int	4	A binary code to differentiate the file from which a given award originates; coded 1 if the award is present in Subaward file; coded 0 otherwise
7	Unique Award Number	unique_award_number	varchar	500	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
8	Vendor File	vendor	int	4	A binary code to differentiate the file from which a given award originates; coded 1 if the award is present in Vendor file; coded 0 otherwise

## aux\_emp\_demographics

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Employee Number	emp_number	varchar	200	IRIS-generated unique identifier assigned to all personnel paid by awards
2	Imputed Race Ethnicity Census	imputed_race_ethnicity_census	varchar	200	Race/ethnicity imputed by use of the Python ethnicolr package and census data; 5 categories include: Asian Pacific Islander, Black, Hispanic, White, and unknown
3	Imputed Race Ethnicity Ethnea Broad	imputed_race_ethnicity_ethnea_broad	varchar	200	Narrowly categorized Ethnea-generated imputed ethnicity groups; 7 categories include: African, Asian, European, Hispanic, Middle Eastern, Other, and unknown
4	Imputed Race Ethnicity Ethnea Narrow	imputed_race_ethnicity_ethnea_narrow	varchar	200	Imputed ethnicity generated by the use of the Ethnea method and data; in addition to 28 broad ethnicity groups (e.g., French, Chinese, English, Thai, etc.), this field has as many as 400+ hyphenated categories to represent multiple ethnicities (e.g., Arab-Hispanic, English-German); this field is null if an employee name was neither found nor matched to the source data that IRIS used to impute ethnicity
5	Imputed Race Ethnicity Wiki	imputed_race_ethnicity_wiki	varchar	200	Race/ethnicity imputed by use of the Python ethnicolr package and census data; 13 categories including Asian-GreaterEastAsian, Asian-IndianSubContinent, GreaterAfrican-Muslim, GreaterAfrican-Africans, GreaterEuropean-WestEuropean-French, GreaterEuropean-WestEuropean-Hispanic, etc.
6	Imputed Gender	imputed_gender	varchar	200	Imputed gender (M for male and F for female) through the use of Ethnea and Python gender-guesser library; this field is unknown if an employee name was neither found nor matched to the Ethnea's source data that used to impute gender
7	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
8	Year of Birth Category	yob_category	int	4	To simplify age bands in the field of "yob_range" above, we converted age bands into integers between 1 and 11. If yob_range value is "na" it is coded as 99, and if "masked" it is coded as 98



9	Year of Birth Range	yob_range	varchar	50	The range in which an employee's year of birth falls, e.g., if one's birth year falls between 1999 and 2003, it is coded as "between_1999_and_2003". Employee age is mapped to the unique employee ID found in the Employee File, and due to disclosure risk, 11 age bands are used to aggregate years. Additional values include "na" and "masked." If no information on an employee's birth year was provided by universities, it is coded as "na". Some age bands are also rolled up into one and coded as "masked" due to disclosure risk
---	---------------------	-----------	---------	----	---

## aux\_institution\_fastfacts

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Carnegie Classification	carnegie_classification	varchar	50	Derived from the 2018 Classification Update of the traditional Carnegie Classification Framework for each academic institution; data retrieved Carnegie Classification of Institutions of Higher Education website: <a href="http://carnegieclassifications.iu.edu/">http://carnegieclassifications.iu.edu/</a>
2	Fall Enrollment	fall_enrollment	int	4	The number of students enrolled in courses that are creditable toward a degree, diploma, certificate, or other formal award, or are part of a vocational or occupational program including any students enrolled in off-campus centers; data retrieved from the Integrated Postsecondary Education Data System (IPEDS) Fall Enrollment Survey
3	Federally Financed R&D Expenditures in All Fields	fed_rd_expenditures	int	4	R&D expenditures in all fields, including direct and recovered indirect costs, funded by all agencies of the Federal government; data retrieved from NSF HERD
4	Institution Control	institution_control	int	4	Defined for academic institutions as private or public (not applicable to biomedical institutions); values include 1 (Private) and 0 (Public); data retrieved from NSF HERD
5	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
6	Land Grant	land_grant	int	4	Indicator for whether an institution is a Land Grant institution; values include 1 (institution is a Land Grant institution) and 0 (not a Land Grant institution); data retrieved from IPEDS and verified on <a href="https://nifa.usda.gov/land-grant-colleges-and-universities-partner-website-directory">https://nifa.usda.gov/land-grant-colleges-and-universities-partner-website-directory</a>
7	Region	region	varchar	30	Indication of the institution's location in regional categories (Northeast, Midwest, South, and West) used by the US Census Bureau
8	Distance between medical school and main campus	main_med_distance	int	4	The geographical distance (in miles) of the medical school from the main campus if an IRIS member university has a medical school
9	Medical School	med_school	int	4	Indicator for each institution having a medical school included as part of its reporting unit; Values include 1 (has medical school) and 0 (does not include medical school); data retrieved from NSF HERD
10	Number of Doctorate Recipients	number_doc_recipients	int	4	All earned doctorates granted by universities; data retrieved from the NSF Survey of Earned Doctorates (SED)
11	Number of Graduate Students	number_grad_students	int	4	The number of graduate students enrolled in GSS-eligible science, engineering, and health (SEH) units in the fall of the data collection year; data retrieved from the NSF-NIH Survey of Graduate Students & Postdoctorates in Science and Engineering (GSS)
12	Number of Principal Investigators	number_pis	int	4	Personnel paid from the R&D salaries, wages and fringe benefits reported on the survey (NSF Research and Development Expenditures at Universities and Colleges/Higher Education Research and Development Survey), and designated by the institution to direct the R&D project or program and be responsible for the scientific and technical direction of the project; Co-investigators (co-PIs) may be designated for this role and are also included. Missing data for this question were not imputed, therefore aggregate totals represent an undercount; data retrieved from NSF HERD
13	Number of Postdocs	number_post_docs	int	4	Personnel defined as postdocs, namely recent doctorate recipients with limited-term appointments primarily intended for training under the supervision of a senior scholar. Data retrieved from the NSF-NIH Survey of Graduate Students & Postdoctorates in Science and Engineering (GSS)
14	Total R&D Expenditures in All Fields	total_rd_expenditures	int	4	R&D expenditures from the institution's current operating funds that were separately accounted for, including expenditures for organized research as defined by 2 CFR 220 Part 200 Appendix III and expenditures from funds designated for research. Expenditures came from internal or external funding and included recovered and unrecovered indirect costs. Funds passed through to subrecipient organizations were also included. R&D was excluded if it was conducted by university faculty or staff at outside institutions and was not accounted for in the reporting institution's financial records. Data retrieved from NSF HERD

15	Year	year	int	4	Year (ranging between 2010 and 2017) is defined in two ways: 1) Academic year: doctorate recipients, fall enrollment, and other personnel-related data; 2) Fiscal year: R&D expenditures as defined in the data source, the NSF Higher Education R&D Survey (NSF HERD)
----	------	------	-----	---	--

## aux\_object\_code

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
2	Object Code	object_code	varchar	50	Internal object code or other expense type category assigned to a transaction to identify payment purposes or resources
3	Object Code Description	object_code_desc	varchar	500	Description of internal object code or other expense type category assigned to a transaction; maps to object code

## aux\_suborg

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Campus ID	campus_id	varchar	37	IRIS-generated identifier assigned to each campus for de-identification purposes. Each ID is a combination of institution_id and a serial number helpful to identify the campus to which a given award is made and by which the award is being administered
2	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3	Main Campus Flag	main_campus	int	4	A binary value to indicate the main campus of each member university; valued 1 if a given campus is the main
4	Sub-organization Unit	sub_org_unit	varchar	62	IRIS-generated identifier assigned to sub-organization units to which each funded project is assigned, such as a
5	Sub-organization Unit Name	sub_org_unit_name	varchar	200	Sub-organizational unit name that maps to sub-organizational unit code, e.g., the college of natural sciences, the
6	Standardized Sub-organization Unit	sub_org_unit_std	varchar	100	Standardized sub-organization unit name with 18 categories (e.g., arts & sciences, health sciences, administration,

## link\_doe\_xwalk

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Award ID	award_id	varchar	100	The DOE unique award identifier assigned to each DOE award that is matched to UMETRICS award number
2	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3	Unique Award Number	unique_award_number	varchar	400	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
4	Match Source	match_source	varchar	100	Federal agency award data source used for award number matching

## link\_dod\_xwalk

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Award ID	award_id	varchar	100	The DOD unique award identifier assigned to each DOD award that is matched to UMETRICS award number
2 Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3 Unique Award Number	unique_award_number	varchar	400	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
4 Match Source	match_source	varchar	100	Federal agency award data source used for award number matching

## link\_nasa\_xwalk

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Award ID	award_id	varchar	100	The NASA unique award identifier assigned to each NASA award that is matched to UMETRICS award number
2 Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3 Unique Award Number	unique_award_number	varchar	400	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
4 Match Source	match_source	varchar	100	Federal agency award data source used for award number matching

## link\_nih\_xwalk

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Core Project Number	core_project_num	varchar	200	The NIH core project number assigned to each NIH-funded project that is matched to the core project part of UMETRICS award number
2 Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3 Unique Award Number	unique_award_number	varchar	200	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)

## link\_nsf\_xwalk

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Award ID	award_id	varchar	100	The NSF unique award identifier assigned to each NSF award that is matched to UMETRICS award number
2 Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers

3	Unique Award Number	unique_award_number	varchar	400	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
---	---------------------	---------------------	---------	-----	---

## link\_usda\_xwalk

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Award ID	award_id	varchar	100	The USDA unique award identifier assigned to each USDA award that is matched to UMETRICS award number
2	Project Number	project_number	varchar	100	The USDA unique project number assigned to each USDA award that is matched to UMETRICS award number
3	Accession Number	accession_number	varchar	100	The USDA unique accession number assigned to each USDA award
4	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
5	Unique Award Number	unique_award_number	varchar	400	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
6	Field Used for Matching	field_used_to_match	varchar	100	Identification of which USDA identifier (award id or project number) was used to link to the unique award number in UMETRICS

## link\_nih\_pub\_xwalk

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Core Project Number	core_project_num	varchar	200	The NIH core project number assigned to each NIH-funded project that is matched to the core project part of UMETRICS award number
2	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
3	PMID	pmid	int	4	PubMed unique identifier assigned by the NIH National Library of Medicine to papers indexed in PubMed (index of abstracts). The number is 1- to 8- digits with no leading zeros
4	Unique Award Number	unique_award_number	varchar	200	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)

## link\_nsf\_pub\_xwalk

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
2	Award ID	award_id	varchar	10	NSF assigned award number (a seven digit number)

3	Unique Award Number	unique_award_number	varchar	100	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)
4	Unique Publication ID	unique_pub_id	varchar	50	IRIS-generated identifier assigned to each NSF-funded publication. Each ID is a combination of NSF award ID and a serial number helpful to identify the award to which a given publication is funded; e.g., the two publications, xxxxxx-1 and xxxxxx-2, are both funded by the same NSF award
5	Unique DOI ID	unique_doi_id	int	4	IRIS-generated unique identifier assigned to each NSF-funded publication based on its unique DOI

## link\_pi\_xwalk

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Agency	agency	varchar	3	Sponsoring agency of a given award associated with the 'unique_award_number' field; values take either 'nih' or 'nsf'
2	Award ID	award_id	varchar	200	Unique award number assigned by an agency, corresponding to the Award ID field for DOD, DOE, NASA, and NSF. For NIH this corresponds to Core Project Number and for USA it corresponds to Accession Number
3	Employee Number	emp_number	varchar	200	IRIS-generated unique identifier assigned to all personnel paid by awards
4	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
5	Unique Award Number	unique_award_number	varchar	200	University-generated unique identifier specifying an award and its funding source, made up of the 5-digit funding source code (e.g., CFDA number) and an award identifier. Award identifier may include the awarding agency's federal award ID (e.g., federal grant number, contract number, or loan number) or an internal award ID for non-federal awards. Values may include a space or dash in between them: e.g., "10.310 2010-12345-54321" (USDA example), "47.050 1234567" (NSF example), "93.865 2-R01-DK-012345-15-S1" (NIH example)

## link\_doe\_source

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the award
2	Action Type	action_type	varchar	200	Type of funded project proposal, e.g., new, revision, adjustment, closeout, etc.
3	Award ID	award_id	varchar	200	The agency assigned award number
4	Award Type	award_type	varchar	100	Type of award, e.g., Grant, Cooperative agreement, Interagency agreement, etc.
5	DOE Data Source	doe_data_source	varchar	50	Name of data source from which IRIS collected agency award data; Three sources include: 1) DOE PAMS, 2) USASpending Assistance, 3) USASpending Contract; See documentation for URLs of each source
6	Modification Number	modification_number	varchar	50	The agency assigned award modification number if applicable
7	Project End Date	project_end_date	date	3	The end date of a project
8	Project Start Date	project_start_date	date	3	The start date of a project
9	Status	status	varchar	100	Status of the award, e.g., active or inactive
10	Title	title	varchar	1000	Descriptive title of the project
11	Total Obligated Amount	total_obligated_amount	numeric	9	Total obligated amount of the award

## link\_dod\_source

	Field Name	Column Name	Data Type	Set Length	Field Definition
--	------------	-------------	-----------	------------	------------------

1	Abstract	abstract	varchar	8000	Abstract of the award
2	Action Type	action_type	varchar	200	Type of funded project proposal, e.g., new, revision, adjustment, closeout, etc.
3	Award ID	award_id	varchar	200	The agency assigned award number
4	Award Type	award_type	varchar	100	Type of award, e.g., Grant, Cooperative agreement, Interagency agreement, etc.
5	DOD Data Source	dod_data_source	varchar	50	Name of data source from which IRIS collected agency award data; Three sources include: 1) DOD CDMRP, 2) DOD Grantawards, 3) USASpending; See documentation for URLs of each source
6	Fiscal Year	fiscal_year	int	4	Fiscal year in which a given project is funded
7	Modification Number	modification_number	varchar	50	The agency assigned award modification number if applicable
8	Project End Date	project_end_date	date	3	The end date of a project
9	Project Start Date	project_start_date	date	3	The start date of a project
10	Status	status	int	4	Status of the award, e.g., active or inactive
11	Title	title	varchar	1000	Descriptive title of the project
12	Total Obligated Amount	total_obligated_amount	numeric	9	Total obligated amount of the award

## link\_nasa\_source

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the award
2	Award ID	award_id	varchar	200	The agency assigned award number
3	CFDA	cfda	varchar	200	A CFDA (Catalog of Federal Domestic Assistance) number assigned and used by NASA to identify and sort federal financial assistance programs
4	Modification Number	modification_number	varchar	50	Indication of modification(s) of the award if applicable
5	NASA Data Source	nasa_data_source	varchar	50	Name of data source from which IRIS collected agency award data; Three sources include: 1) FederalReporter, 2) NASA API, 3) USASpending; See documentation for URLs of each source
6	Project End Date	project_end_date	date	3	The end date of a project
7	Project Start Date	project_start_date	date	3	The start date of a project
8	Title	title	varchar	8000	Descriptive title of the project
9	Total Obligated Amount	total_obligated_amount	numeric	9	Total obligated amount of the award

## link\_nih\_source

NIH EXPORTER Data Dictionary also available from the following URL: <https://report.nih.gov/exporter-data-dictionary>

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the funded project
2	Activity	activity	varchar	50	A 3-character code identifying the grant, contract, or intramural activity through which a project is supported. Within each funding mechanism, NIH uses 3-character activity codes (e.g., F32, K08, P01, R01, T32, etc.) to differentiate the wide variety of research-related programs NIH supports
3	Administrator IC	administering_ic	varchar	50	Administering Institute or Center. A two-character code to designate the agency, NIH Institute, or Center administering the grant
4	Application ID	application_id	int	4	A unique identifier of the project record

5	Application Type	application_type	varchar	50	A one-digit code to identify the type of application funded: 1) New Application; 2) Competing continuation; 3) Application for additional support; 4) Competing extension for an R37 award or first non-competing year of a Fast Track SBIR/STTR award; 5) Non-competing continuation; 7) Change of grantee institution; 9) Change of NIH awarding Institute or Division
6	ARRA Funded	arra_funded	varchar	50	"Y" indicates a project supported by funds appropriated through the American Recovery and Reinvestment Act of 2009
7	Award Notice Date	award_notice_date	date	3	Award notice date or Notice of Grant Award (NGA) is a legally binding document stating the government has obligated funds and which defines the period of support and the terms and conditions of award
8	Budget End	budget_end	date	3	The date when a project's funding for a particular fiscal year ends
9	Budget Start	budget_start	date	3	The date when a project's funding for a particular fiscal year begins
10	CFDA Code	cfda_code	varchar	50	A CFDA (Catalog of Federal Domestic Assistance) number used to identify and sort federal financial assistance programs
11	Core Project Number	core_project_num	varchar	50	The NIH core project number assigned to each NIH-funded project that is matched to the core project part of UMETRICS award number
12	COVID-research Grant Flag	covid_grant_flag	int	4	A binary indicator identifying a COVID-research grant; coded 1 if the award is COVID-research grant; coded 0 otherwise
13	Direct Cost	direct_cost_amt	numeric	9	Total direct cost funding for a project for a given fiscal year. Available only for NIH awards funded in FY 2012 onward and not available for SBIR/STTR awards
14	ED Inst Type	ed_inst_type	varbinary	50	Institution type
15	FOA Number	foa_number	varchar	50	The number of the funding opportunity announcement, if any, under which the project application was solicited. Funding opportunity announcements may be categorized as program announcements, requests for applications, notices of funding availability, solicitations, or other names depending on the agency and type of program. Funding opportunity announcements can be found at Grants.gov/FIND and in the NIH Guide for Grants and Contracts
16	Full Project Number	full_project_num	varchar	50	Commonly referred to as a grant number, intramural project, or contract number. For grants, this unique identification number is composed of the type code, activity code, Institute/Center code, serial number, support year, and (optional) a suffix code to designate amended applications and supplements
17	Funding ICs	funding_ics	varchar	50	The NIH Institute or Center(s) providing funding for a project are designated by their acronyms (see Institute/Center acronyms). Each funding IC is followed by a colon (:) and the amount of funding provided for the fiscal year by that IC. Multiple ICs are separated by semicolons (;). Project funding information is available only for NIH projects awarded in FY 2008 and later fiscal years
18	FY	fy	int	4	The fiscal year appropriation from which project funds were obligated
19	IC Name	ic_name	varchar	100	Full name of the administering agency, Institute, or Center
20	Standardized IC Name	ic_name_standardized	varchar	100	Standardized name of the administering agency, Institute, or Center
21	Indirect Cost	indirect_cost_amt	numeric	9	Total indirect cost funding for a project for a given fiscal year. Available only for NIH awards funded in FY 2012 and onward and not available for SBIR/STTR awards
22	NIH Spending CATS	nih_spending_cats	varchar	500	Congressionally-mandated reporting categories into which NIH projects are categorized. Available for fiscal years 2008 and later. Each project's spending category designations for each fiscal year are made available the following year as part of the next President's Budget request. See the Research, Condition, and Disease Categorization System for more information on the categorization process
23	Org Dept	org_dept	varchar	50	The departmental affiliation of the contact principal investigator for a project, using a standardized categorization of departments. Names are available only for medical school departments
24	Org District	org_district	varchar	50	The congressional district in which the business office of the grantee organization or contractor is located. Note that this may be different from the research performance site
25	Org FIPS	org_fips	varchar	50	The country code of the grantee organization or contractor as defined in the Federal Information Processing Standard
26	PHR	phr	varchar	2000	Submitted as part of a grant application, this statement articulates a project's potential to improve public health
27	PI IDs	pi_ids	varchar	50	A unique identifier for each of the project Principal Investigators. Each PI in the RePORTER database has a unique identifier that is constant from project to project and year to year, but changes may be observed for investigators that have had multiple accounts in the past, particularly for those associated with contracts or sub-projects

28	Project End	project_end	date	3	The current end date of the project, including any future years for which commitments have been made. For subprojects of a multi-project grant, this is the end date of the parent award. Upon competitive renewal of a grant, the project end date is extended by the length of the renewal award
29	Project Start	project_start	date	3	The start date of a project. For subprojects of a multi-project grant, this is the start date of the parent award
30	Project Terms	project_terms	varchar	3000	Thesaurus terms assigned by NIH CRISP indexers, only applicable to projects funded prior to the fiscal year 2008
31	Project Title	project_title	varchar	500	Title of the funded grant, contract, or intramural (sub)project
32	Serial Number	serial_number	varchar	50	A six-digit number assigned in serial number order within each administering organization
33	Study Section	study_section	varchar	50	A designator of the legislatively-mandated panel of subject matter experts that reviewed the research grant application for scientific and technical merit
34	Study Section Name	study_section_name	varchar	100	The full name of a regular standing Study Section that reviewed the research grant application for scientific and technical merit. Applications reviewed by panels other than regular standing study sections are designated by "Special Emphasis Panel"
35	Sub Project ID	subproject_id	int	50	A unique numeric designation assigned to subprojects of a "parent" multi-project research grant
36	Suffix	suffix	varchar	50	A suffix to the grant application number that includes the letter "A" and a serial number to identify an amended version of an original application and/or the letter "S" and serial number indicating a supplement to the project
37	Support Year	support_year	int	4	The year of support for a project, as shown in the full project number. For example, a project with number 5R01GM0123456-04 is in its fourth year of support
38	Total Cost	total_cost	numeric	9	Total project funding from all NIH Institute and Centers for a given fiscal year. Costs are available only for: 1) NIH and CDC grant awards (only the parent record of multi-project grants) funded in FY 2000 and later fiscal years; 2) NIH intramural projects (activity codes beginning with "Z") in FY 2007 and later fiscal years; 3) NIH contracts (activity codes beginning with "N") in FY 2007 and later fiscal years. For multi-project grants, Total Cost includes funding for all of the constituent subprojects. This field will be blank on subproject records; the total cost of each subproject is found in Total_Cost_Sub_Project (FY 2000 and later fiscal years only)
39	Total Cost Sub Project	total_cost_sub_project	numeric	9	Applies to subproject records only. Total funding for a subproject from all NIH Institute and Centers for a given fiscal year. Costs are available only for NIH awards funded in FY 2000 and later

## link\_nsf\_source

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the award
2	ARRA Amount	arra_amount	varchar	50	Amount of funding obligated designated as ARRA funding
3	Award Amount	award_amount	numeric	9	The amount obligated to date for the project
4	Award Effective Date	award_effective_date	date	3	Effective date of the award
5	Award Expiration Date	award_expiration_date	date	3	The date on which the award expires
6	Award ID	award_id	varchar	20	The agency assigned award number (a seven digit number)
7	Award Instrument	award_instrument	varchar	100	Type of Award
8	Award Instrument Code	award_instrument_code	varchar	100	Code associated with type of award
9	Award Title	award_title	varchar	500	Descriptive title of the project
10	COVID-research Grant Flag	covid_grant_flag	int	4	A binary indicator identifying a COVID-research grant; coded 1 if the award is COVID-research grant; coded 0 otherwise
11	Directorate	directorate	varchar	200	Department / Directorate of NSF funding the award
12	Directorate Code	directorate_code	varchar	50	Code (acronym) for NSF Directorate, e.g., EHR for Directorate for Education & Human Resources, BIO for Directorate for Biological Sciences
13	Division	division	varchar	200	Division of NSF funding the award
14	Standardized Directorate Name	directorate_name_standardized	varchar	200	Cleaned and standardized name of the Department / Directorate of NSF funding the award
15	Standardized Division Name	division_name_standardized	varchar	200	Cleaned and standardized name of the Division of NSF funding the award



## link\_usda\_source

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Abstract	abstract	varchar	8000	Abstract of the award
2 Accession Number	accession_number	varchar	200	One of the two project identifiers for each funded project; Accession Numbers begin with zero (0) and consist of seven (7) digits, e.g., accession number looks like 0134036
3 Award ID	award_id	varchar	200	The agency assigned award identifier
4 Cumulative Award Amount	cumulative_award_amount	numeric	9	Cumulative award amount from the start date of a project to a given fiscal year
5 Funding Mechanism	funding_mechanism	varchar	200	Type of funding mechanism, including: NIFA Formula, NIFA Non Formula, Other USDA Funding, State Funding, etc.
6 Funding Source	funding_source	varchar	200	Type of funding source, including: Evans-Allen, Hatch, Hatch/Multi-State, McIntire-Stennis, Non Formula, State, etc.
7 Methods	methods	varchar	8000	Description of project methods
8 Multistate Project Number	multistate_project_number	varchar	200	The Multistate (Regional) Project No. field (RL) contains a regional abbreviation and a numeric designation separated by a hyphen (-). Only those projects which contribute to a particular Multistate Project will have data in this field. The possible codes and their translations are: IR = Interregional; NC = North Central Region; NE = Northeastern Region; NRSP = National Research Support Project; S = Southern Region, and; W = Western Region
9 Nontechnical Summary	nontechnical_summary	varchar	8000	Nontechnical summary of the project
10 Objectives	objectives	varchar	8000	Objectives of the project
11 Project End Date	project_end_date	date	3	The end date of a project
12 Project Number	project_number	varchar	200	Project number of the awarded project
13 Project Start Date	project_start_date	date	3	The start date of a project
14 Project Status	project_status	varchar	200	The identification of the status of project: active, extended, new, pending, revised, or terminated
15 Project Title	project_title	varchar	1000	Descriptive title of the project
16 Sponsoring Agency	sponsoring_agency	varchar	200	The name of sponsoring institution. These include: State Agricultural Experiment Station; Forest Service/USDA; National Institute of Food and Agriculture; Other Cooperating Institutions; Economic Research Service/USDA; Cooperating Schools of Veterinary Medicine; Rural Business-Cooperative Service; Agricultural Research Service/USDA
17 Total Amount of All Funds	total_all_funds	numeric	9	Total obligated amount of the award

## link\_nih\_pub\_source

Field Name	Column Name	Data Type	Set Length	Field Definition
1 Country	country	varchar	200	The name of the country where publication occurred
2 ISSN	issn	varchar	20	The International Standard Serial Number, an eight-character value that uniquely identifies the journal
3 Journal Issue	journal_issue	varchar	200	An issue, part, or supplement of the journal in which the article was published
4 Journal Title	journal_title	varchar	1000	Full journal title, taken from the NIH National Library of Medicine's cataloging data
5 Journal Title Abbreviation	journal_title_abbr	varchar	500	Standard abbreviation for the title of the journal in which the article appeared
6 Journal Volume	journal_volume	varchar	200	Volume number of the journal in which the article was published
7 Language	lang	varchar	20	Three-letter abbreviation representing the language(s) in which an article was published. List of abbreviations is available at: <a href="https://www.nlm.nih.gov/bsd/language_table.html">https://www.nlm.nih.gov/bsd/language_table.html</a>
8 Page Number	page_number	varchar	500	Pages for the article, including document numbers for electronic articles
9 PMCID	pmc_id	varchar	20	A unique identifier for the article in PubMed Central (index of full-text papers). The PMCID is assigned to each full-text paper in PubMed Central by the National Library of Medicine
10 PMID	pmid	int	4	PubMed unique identifier assigned by the NIH National Library of Medicine to papers indexed in PubMed (index of abstracts). The number is 1- to 8- digits with no leading zeros; this is the field one should use to retrieve publication details when using the award-level NIH - publication - UMETRICS crosswalk

11	Publication Date	pub_date	varchar	50	Date on which the issue of the journal was published. The standardized format includes a 4-digit year, a 3-character abbreviated month, and a 1 or 2-digit day, but the data are taken as published in the journal issue so not every record contains all elements
12	Publication Title	pub_title	varchar	8000	Title of the article; if originally published in a non-English language this is a translation enclosed in brackets
13	Publication Year	pub_year	int	4	Year of publication, from pub_date

## link\_nsf\_pub\_source

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Award ID	award_id	varchar	10	NSF assigned award number (a seven digit number)
2	CrossRef ID	crossref_id	varchar	100	Publication ID assigned by CrossRef
3	CrossRef Member ID	crossref_member_id	int	4	Crossref member ID; see <a href="https://www.crossref.org/reports/members-with-open-references/">https://www.crossref.org/reports/members-with-open-references/</a>
4	Date Created	created_date	date	3	Date the DOI was minted
5	DOI	doi	varchar	100	The Digital Object Identifier (DOI), a unique alphanumeric string beginning with '10.####/text' assigned by a registration agency to provide a persistent link to the location of a publication or other digital object
6	ISBN	isbn	varchar	100	International Standard Book Number (ISBN), a unique numeric identifier
7	ISSN	issn	varchar	100	International Standard Serial Number (ISSN), an eight-digit serial number used to uniquely identify a serial publication
8	Issue	issue	varchar	100	The issue number in which an article is published
9	Journal Name	journal_name	varchar	1000	Name of the journal
10	Language	lang	varchar	50	Language of the publication
11	Number of Pages	page_ct	varchar	50	Pages for the article
12	Publication Type	pub_type	varchar	50	Type of publication, e.g., journal-article, book, book chapter, paper-conference, etc.
13	Published in Print	pub_print	date	3	Date published in print
14	Published Online	pub_online	date	3	Date published online
15	Publisher	publisher	varchar	200	Publisher name
16	Reference Count	ref_ct	int	4	Number of references in the publication
17	Referenced By Count	referenced_by_ct	int	4	Number of times that a publication was cited
18	Short Journal Name	short_journal_name	varchar	200	Abbreviated name of the journal (e.g., Am. J. Hum. Biol. for the American Journal of Human Biology)
19	Subject	subject	varchar	500	Subject fields assigned to the publication; multiple subjects are combined with ' '
20	Title	title	varchar	4000	Title of the publication
21	Unique DOI ID	unique_doi_id	int	4	IRIS-generated unique identifier assigned to each NSF-funded publication based on its unique DOI.
22	Unique Publication ID	unique_pub_id	varchar	50	IRIS-generated identifier assigned to each NSF-funded publication. Each ID is a combination of NSF award ID and a serial number helpful to identify the award to which a given publication is funded; e.g., the two publications, xxxxxx-1 and xxxxxx-2, are both funded by the same NSF award.
23	URL	url	varchar	200	URL of the publication from the DOI resolver
24	Volume	volume	varchar	50	The volume number of a published journal, or the number of a printed volume for a book or conference proceedings

## link\_nih\_lda\_abstract\_topics

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the award
2	Core Project Number	core_project_num	varchar	50	The NIH core project number assigned to each NIH-funded project that is matched to the core project part of UMETRICS award number

3	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
4	Topic 1	topic1	int	4	Topic identifier corresponding to the highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
5	Topic 1 probability	topic1_prob	numeric	9	The probability of Topic 1 for the given NIH abstract.
6	Topic 2	topic2	int	4	Topic identifier corresponding to the second highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
7	Topic 2 probability	topic2_prob	numeric	9	The probability of Topic 2 for the given NIH abstract.
8	Topic 3	topic3	int	4	Topic identifier corresponding to the third highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
9	Topic 3 probability	topic3_prob	numeric	9	The probability of Topic 3 for the given NIH abstract.
10	Topic 4	topic4	int	4	Topic identifier corresponding to the fourth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
11	Topic 4 probability	topic4_prob	numeric	9	The probability of Topic 4 for the given NIH abstract.
12	Topic 5	topic5	int	4	Topic identifier corresponding to the fifth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
13	Topic 5 probability	topic5_prob	numeric	9	The probability of Topic 5 for the given NIH abstract.
14	Topic 6	topic6	int	4	Topic identifier corresponding to the sixth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
15	Topic 6 probability	topic6_prob	numeric	9	The probability of Topic 6 for the given NIH abstract.
16	Topic 7	topic7	int	4	Topic identifier corresponding to the seventh highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
17	Topic 7 probability	topic7_prob	numeric	9	The probability of Topic 7 for the given NIH abstract.
18	Topic 8	topic8	int	4	Topic identifier corresponding to the eighth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
19	Topic 8 probability	topic8_prob	numeric	9	The probability of Topic 8 for the given NIH abstract.
20	Topic 9	topic9	int	4	Topic identifier corresponding to the ninth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
21	Topic 9 probability	topic9_prob	numeric	9	The probability of Topic 9 for the given NIH abstract.
22	Topic 10	topic10	int	4	Topic identifier corresponding to the tenth highest topic probability for a given abstract, as determined during LDA topic modeling using NIH abstracts.
23	Topic 10 probability	topic10_prob	numeric	9	The probability of Topic 10 for the given NIH abstract.

## link\_nsf\_lda\_abstract\_topics

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Abstract	abstract	varchar	8000	Abstract of the award
2	Award ID	award_id	varchar	20	The agency assigned award number (a seven digit number)
3	Institution ID	institution_id	int	4	IRIS-generated unique identifier assigned to each IRIS member university for de-identification purposes. Values are four or five digit numbers
4	Topic 1	topic1	int	4	Topic identifier corresponding to the highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
5	Topic 1 probability	topic1_prob	numeric	9	The probability of Topic 1 for the given NSF abstract.
6	Topic 2	topic2	int	4	Topic identifier corresponding to the second highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
7	Topic 2 probability	topic2_prob	numeric	9	The probability of Topic 2 for the given NSF abstract.
8	Topic 3	topic3	int	4	Topic identifier corresponding to the third highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.

9	Topic 3 probability	topic3_prob	numeric	9	The probability of Topic 3 for the given NSF abstract.
10	Topic 4	topic4	int	4	Topic identifier corresponding to the fourth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
11	Topic 4 probability	topic4_prob	numeric	9	The probability of Topic 4 for the given NSF abstract.
12	Topic 5	topic5	int	4	Topic identifier corresponding to the fifth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
13	Topic 5 probability	topic5_prob	numeric	9	The probability of Topic 5 for the given NSF abstract.
14	Topic 6	topic6	int	4	Topic identifier corresponding to the sixth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
15	Topic 6 probability	topic6_prob	numeric	9	The probability of Topic 6 for the given NSF abstract.
16	Topic 7	topic7	int	4	Topic identifier corresponding to the seventh highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
17	Topic 7 probability	topic7_prob	numeric	9	The probability of Topic 7 for the given NSF abstract.
18	Topic 8	topic8	int	4	Topic identifier corresponding to the eighth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
19	Topic 8 probability	topic8_prob	numeric	9	The probability of Topic 8 for the given NSF abstract.
20	Topic 9	topic9	int	4	Topic identifier corresponding to the ninth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
21	Topic 9 probability	topic9_prob	numeric	9	The probability of Topic 9 for the given NSF abstract.
22	Topic 10	topic10	int	4	Topic identifier corresponding to the tenth highest topic probability for a given abstract, as determined during LDA topic modeling using NSF abstracts.
23	Topic 10 probability	topic10_prob	numeric	9	The probability of Topic 10 for the given NSF abstract.

## link\_nih\_lda\_abstract\_topic\_terms

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Term 1	term1	varchar	50	Term corresponding to the highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
2	Term 1 probability	term1_prob	numeric	9	The probability of Term 1 for the given topic.
3	Term 2	term2	varchar	50	Term corresponding to the second highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
4	Term 2 probability	term2_prob	numeric	9	The probability of Term 2 for the given topic.
5	Term 3	term3	varchar	50	Term corresponding to the third highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
6	Term 3 probability	term3_prob	numeric	9	The probability of Term 3 for the given topic.
7	Term 4	term 4	varchar	50	Term corresponding to the fourth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
8	Term 4 probability	term4_prob	numeric	9	The probability of Term 4 for the given topic.
9	Term 5	term5	varchar	50	Term corresponding to the fifth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
10	Term 5 probability	term5_prob	numeric	9	The probability of Term 5 for the given topic.
11	Term 6	term6	varchar	50	Term corresponding to the sixth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
12	Term 6 probability	term6_prob	numeric	9	The probability of Term 6 for the given topic.
13	Term 7	term7	varchar	50	Term corresponding to the seventh highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
14	Term 7 probability	term7_prob	numeric	9	The probability of Term 7 for the given topic.

15	Term 8	term8	varchar	50	Term corresponding to the eighth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
16	Term 8 probability	term8_prob	numeric	9	The probability of Term 8 for the given topic.
17	Term 9	term9	varchar	50	Term corresponding to the ninth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
18	Term 9 probability	term9_prob	numeric	9	The probability of Term 9 for the given topic.
19	Term 10	term10	varchar	50	Term corresponding to the tenth highest term probability for a given topic, as determined during LDA topic modeling using NIH abstracts.
20	Term 10 probability	term10_prob	numeric	9	The probability of Term 10 for the given topic.
21	Topic ID	topic_id	int	4	Integer identifying the topic, as determined during LDA topic modeling using NIH abstracts. Used to link to the link_nih_lda_abstract_topics table.

## link\_nsf\_lda\_abstract\_topic\_terms

	Field Name	Column Name	Data Type	Set Length	Field Definition
1	Term 1	term1	varchar	50	Term corresponding to the highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
2	Term 1 probability	term1_prob	numeric	9	The probability of Term 1 for the given topic.
3	Term 2	term2	varchar	50	Term corresponding to the second highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
4	Term 2 probability	term2_prob	numeric	9	The probability of Term 2 for the given topic.
5	Term 3	term3	varchar	50	Term corresponding to the third highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
6	Term 3 probability	term3_prob	numeric	9	The probability of Term 3 for the given topic.
7	Term 4	term4	varchar	50	Term corresponding to the fourth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
8	Term 4 probability	term4_prob	numeric	9	The probability of Term 4 for the given topic.
9	Term 5	term5	varchar	50	Term corresponding to the fifth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
10	Term 5 probability	term5_prob	numeric	9	The probability of Term 5 for the given topic.
11	Term 6	term6	varchar	50	Term corresponding to the sixth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
12	Term 6 probability	term6_prob	numeric	9	The probability of Term 6 for the given topic.
13	Term 7	term7	varchar	50	Term corresponding to the seventh highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
14	Term 7 probability	term7_prob	numeric	9	The probability of Term 7 for the given topic.
15	Term 8	term8	varchar	50	Term corresponding to the eighth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
16	Term 8 probability	term8_prob	numeric	9	The probability of Term 8 for the given topic.
17	Term 9	term9	varchar	50	Term corresponding to the ninth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
18	Term 9 probability	term9_prob	numeric	9	The probability of Term 9 for the given topic.
19	Term 10	term4	varchar	50	Term corresponding to the tenth highest term probability for a given topic, as determined during LDA topic modeling using NSF abstracts.
20	Term 10 probability	term10_prob	numeric	9	The probability of Term 10 for the given topic.
21	Topic ID	topic_id	int	4	Integer identifying the topic, as determined during LDA topic modeling using NSF abstracts. Used to link to the link_nsf_lda_abstract_topics table.