

PKG2020S4 (1781-Dec. 2020) Database

Description

The database recovery process was successfully tested under ubuntu 18.04.3, using a mysql 5.7.28.

Each of the 27 compressed files corresponding to one table in the database PKG2020S4. Please do the following steps to restore the PKG2020S4 to your MySQL database.

1. Download all the 27 compressed files as well as their MD5 files in the folder “PKG2020S4_MySQL”;

2. Verify each compressed file with their corresponding MD5 file. For example, we can use the following command to verify if the file “PKG2020S4_A01_Articles.sql.gz” download without any damage:

```
md5sum -c PKG2020S4_A01_Articles.sql.gz.md5sum
```

3. Create a new Database in your MySQL server, and make sure the new database “Charset” is set to: utf8mb4, and “Order rule” is set to: utf8mb4_bin.

4. Next, you can inject every table into the target database using the command like:

```
gunzip < PKG2020S4_A01_Articles.sql.gz | mysql -uusername -ppassword  
destinationDatabaseName
```

The above command will import the table A01_Articles.sql.gz into destination database. Similarly,

The tables starting with the beginning of A are the original data tables of PubMed. The tables with the beginning of B are the tables associated with external data sources to provide a support for disambiguation work. The table with the beginning of C is a table synthesized by extracting key information from the tables A and B for subsequent statistics and calculations.

The original document of each field description of PubMed can be found at:

https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html

1 A01_Articles

Specific information for each article

Column Name	Description
id	Automatically incremented identifier
PMID	Unique ID assigned by PubMed to identify PubMed articles.
PMID_Version	PMID version number, which is the record corresponding to all PMID versions in this table. That is, if the PMID version number is different, there are different records in this table, and the id field of each record is different. In other related article tables (such as C01_Article_simple), only the record

	with the highest version number is kept.
MedlineCitation_Owner	Organization responsible for creating and verifying citations, including: NLM, NASA, PIP, KIE, HSR, HMD, SIS, NOTNLM
MedlineCitation_Status	It is the stage of the article. There are seven possible values: Completed In-Process PubMed-not-MEDLINE In-Data-Review Publisher MEDLINE OLDMEDLINE
Journal_JournalIssue_PubDate_Year	Publication year of the current article.
Journal_JournalIssue_PubDate_MedlineDate	Publication year of the current article. If <i>Journal_JournalIssue_PubDate_Year</i> is null, the publication year of the current article is recorded in this field

2 A02_AuthorList

Specific information for each author

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Au_Order	Author order of the current authors in the author list of current articles.
LastName	Last name of the current author.
ForeName	Current author's name excluding the last name and suffix
LastNameForeName	MD5 code generated by the LastName and the first letter of ForeName
AuthorNum	Co-author number of the current article
Vetle_aid	Unique author ID allocated by Vetle.
ORCID	Unique researcher ID that distinguishes the researcher from others allocated by ORCID.
StrongID	Unique author ID, we assign the author ID to each author after disambiguation according to strong features
S2ID	Unique author ID allocated by Semantic Scholar
PIID	If the current author is one of the Principal Investigators of NIH, there is PIID in this field. It is 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.
aid	Unique author ID (The final author disambiguation result. S2id is the main source, supplemented by other disambiguation results)

3 A03_KeywordList

Article keyword information

keyword information in this table is provided by the data producer

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Keyword	Keywords of the current article

4 A04_Abtract

The abstract of each article

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
AbstractText	The abstract of the current Article
KeywordExtraction	Keyword information extracted by the Rake algorithm, and the weight corresponding to the keyword

5 A05_GrantList

Grants details of each article

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
GrantID	Funding identifiers, including research grant numbers or contract numbers (or both) that are financially supported by the US Public Health Service or any agency of the National Institutes of Health (NIH)
Project_Number	The NIH project number corresponding to the funded article. Each article may be funded by multiple projects. Multiple project numbers are separated by ",".

6 A06_MeshheadingList

Mesh Heading details of each article

Mesh Heading refers to the NLM control vocabulary and medical subject heading (MeSH®), which is used to characterize the content of the articles represented by MEDLINE citations.

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.

DescriptorName	Descriptors extracted from each article
DescriptorName_MajorTopicYN	If the MeSH descriptors assigned to the article is the Key word of the article, the value is Y, otherwise it is N.
DescriptorName_UI	Identify MeSH's unique encoding for each descriptor and qualifier
QualifierName	Qualifiers, including numbers and words

7 A07_SupplMeshList

Supplementary conceptual terms and protocol terms for each article

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
SupplMeshName	Supplementary conceptual term
SupplMeshName_Type	The type of supplementary conceptual term
SupplMeshName_UI	MeSH unique identifier for supplementary protocols and diseases

8 A08_ChemicalList

The chemical substances and registry number covered in each article.

Registry Number refers to a code assigned by Chemical Abstracts Service to a specific chemical substance.

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
RegistryNumber	Unique Identifier to a specific chemical substance assigned by Chemical Abstracts Service
NameOfSubstance	The name of the specific chemical substance

9 A09_CommentsCorrectionsList

Reference information for each article, including the source, type, and PMID of the reference

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
RefSource	Reference sources
RefType	Reference type
RefPMID	The PMID of the reference
RefNote	Corrections to records with incorrect references

10 A10_DatabankList

The search number of the molecular sequence database that appears in the PubMed article. The search number can find the information of the corresponding chemical molecule from the established molecular sequence database, avoiding the use of lengthy molecular formulas and graphics in the article.

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
DataBankName	Name of the molecular sequence database

11 A11_PersonalNameSubjectList

Specific information for each author (this table is not as complete as A02)

12 A12_InvestigatorList

Each article corresponds to the NASA-funded principal investigator (PI) information, and they participated in the discussion and research of the article (but not necessarily the author)

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
AffiliationInfo_Affiliation	The affiliation which the researchers belong to
AffiliationInfo_Identifier	Unique identifier of the affiliation

13 A13_AffiliationList

Extracted affiliation information

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Au_Order	Author order of the current authors in the author list of current articles.
Affiliation	Affiliation string.
Affiliation_order	For authors belonging to multiple affiliations, use numbers to identify different affiliations (in no particular order)

14 A14_ReferenceList

Reference information (we use the information in this table to generate the author's self-cited record during the disambiguation of strong features)

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
RefRank	Reference ranking
RefCitation	Reference sources (journals, year of publication, etc.)
RefArticleId	The PMID of the reference
RefIdType	Reference source database (e.g., PubMed)

15 B01_Descriptor

Specific information of Descriptor Name of each article (used for the classification of articles)

Column Name	Description
DescriptorUI	The unique identifier of the descriptor
DescriptorName	The unique term of the descriptor
DescriptorClass	The corresponding classification of the descriptor (such as subject index, document type)
DateCreated	The date of the recording
DateEstablished	The date available for retrieval
Annotation	Text information corresponding to the descriptors
HistoryNote	History of text messages designed to help online searchers
NLMClassificationNumber	The NLM classification numbers assigned to the index terms
OnlineNote	Text messages designed to help online searchers
PublicMeSHNote	Change history of information
ConsiderAlso	Cross-reference, point to similar descriptors
PreviousIndex	Different usage in Descriptors and SCRs: In Descriptors: Free-text field referring to Descriptors or Descriptor/Qualifier combinations which were used to index the concept in the MEDLINE databases before the Descriptor was created.
TreeNumberList	One or more sets of elements in the descriptor or qualifier record

16 B02_projectlist_NIH

Contains NIH funding information

.1、Download link:

ExPORTER: https://exporter.nih.gov/ExPORTER_Catalog.aspx

CRISP: https://exporter.nih.gov/crisp_catalog.aspx

2、Overview of data consolidation:

In Crisp, there are PI_DEPARTMENT and PI_DEPT_CODE fields, not in Exporter, so we need to merge the data in Crisp into ExPORTER according to APPLICATION_ID, and mark the flag as RC. Since the Crisp table contains data from 1970-2009, ExPORTER contains 1985-2020, so the data from 1985-2009 can be merged, the flags of these tables are listed as RC.

PI_FIRST_NAME, PI_MIDDLE_NAME, PI_LAST_NAME in Crisp can also be merged into ExPORTER table according to the above rules. (The PI_NAMES column in the Exporter table contains the above three columns of information)

Data description:

Column Name	Description
APPLICATION_ID	A unique identifier of the project record in the ExPORTER database.
ACTIVITY	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. A comprehensive list of activity codes for grants and cooperative agreements may be found on the Types of Grant Programs Web page. RePORTER also includes R&D contracts (activity codes beginning with the letter N) and intramural projects (beginning with the letter Z).
ADMINISTERING_IC	Administering Institute or Center - A two-character code to designate the agency, NIH Institute, or Center administering the grant. See Institute/Center code definitions.
APPLICATION_TYPE	A one-digit code to identify the type of application funded:
ARRA_FUNDED	"Y" indicates a project supported by funds appropriated through the American Recovery and Reinvestment Act of 2009.
AWARD_NOTICE_DATE	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.
BUDGET_START	The date when a project's funding for a particular fiscal year begins.
BUDGET_END	The date when a project's funding for a particular fiscal year ends.
CFDA_CODE	Federal programs are assigned a number in the Catalog of Federal Domestic Assistance (CFDA), which is referred to as the "CFDA code." The CFDA database helps the Federal government track all programs it has domestically funded.
CORE_PROJECT_NUMBER	An identifier for each research project, used to associate the project with publication and patent records. This identifier is not specific to any particular year of the project. It consists of the project activity code, administering IC, and serial number (a concatenation of Activity, Administering_IC, and Serial_Number).
ED_INST_TYPE	Generic name for the grouping of components across an institution who has applied for or receives NIH funding. The official name as used by NIH is Major Component Combining Name.
FOA_NUMBER	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 to Grants and Contracts.
FULL_PROJECT_NUMBER	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.
SUBPROJECT_ID	A unique numeric designation assigned to subprojects of a “parent” multi-project research grant.
FUNDING_ICs	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, CDC, FDA, and ACF projects.
FUNDING_MECHANISM	The major mechanism categories used in NIH Budget mechanism tables for the President’s budget. Extramural research awards are divided into three main funding mechanisms: grants, cooperative agreements and contracts. A funding mechanism is the type of funded application or transaction used at the NIH. Within each funding mechanism NIH includes programs. Programs can be further refined by specific activity codes.
FY	The fiscal year appropriation from which project funds were obligated.
IC_NAME	Full name of the administering agency, Institute, or Center.
NIH_SPENDING_CATEGORIES	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.
ORG_CITY	The city in which the business office of the grantee organization or contractor is located. Note that this may be different from the research performance site. For all NIH intramural projects, Bethesda, MD is used.
ORG_COUNTRY	The country in which the business office of the grantee organization or contractor is located. Note that this may be different from the research performance site.
ORG_DEPT	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.
ORG_DISTRICT	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.
ORG_DUNS	This field may contain multiple DUNS Numbers separated by a semi-colon. The Data Universal Numbering System is a unique nine-digit number assigned by Dun and Bradstreet Information Services, recognized as the universal standard for identifying and keeping track of business worldwide.
ORG_FIPS	The country code of the grantee organization or contractor as defined in the Federal Information Processing Standard.

ORG_NAME	The name of the educational institution, research organization, business, or government agency receiving funding for the grant, contract, cooperative agreement, or intramural project.
ORG_STATE	The state in which the business office of the grantee organization or contractor is located. Note that this may be different from the research performance site.
ORG_ZIPCODE	The zip code in which the business office of the grantee organization or contractor is located. Note that this may be different from the research performance site.
PHR	Submitted as part of a grant application, this statement articulates a project's potential to improve public health.
PI_IDS	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.
PI_NAMES	The name(s) of the Principal Investigator(s) designated by the organization to direct the research project.
PROGRAM_OFFICE R_NAME	An Institute staff member who coordinates the substantive aspects of a contract from planning the request for proposal to oversight.
PROJECT_START	The start date of a project. For subprojects of a multi-project grant, this is the start date of the parent award.
PROJECT_END	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.
PROJECT_TERMS	Prior to fiscal year 2008, these were thesaurus terms assigned by NIH CRISP indexers. For projects funded in fiscal year 2008 and later, these are concepts that are mined from the project's title, abstract, and specific aims using an automated text mining tool.
PROJECT_TITLE	Title of the funded grant, contract, or intramural (sub)project.
SERIAL_NUMBER	A six-digit number assigned in serial number order within each administering organization.
STUDY_SECTION	A designator of the legislatively-mandated panel of subject matter experts that reviewed the research grant application for scientific and technical merit.
STUDY_SECTION_N AME	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."
SUFFIX	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. .
SUPPORT_YEAR	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.
DIRECT_COST_AM T	Total direct cost funding for a project from all NIH Institute and Centers for a given fiscal year. Costs are available only for NIH awards funded in FY 2012 onward. Direct cost amounts are not available for SBIR/STTR awards.

INDIRECT_COST_A MT	Total indirect cost funding for a project from all NIH Institute and Centers for a given fiscal year. Costs are available only for NIH awards funded in FY 2012 and onward. Indirect cost amounts are not available for SBIR/STTR awards.
TOTAL_COST	Total project funding from all NIH Institute and Centers for a given fiscal year. Costs are available only for: NIH, CDC, FDA, and ACF grant awards (only the parent record of multi-project grants). NIH intramural projects (activity codes beginning with “Z”) in FY 2007 and later fiscal years. NIH contracts (activity codes beginning with “N”) . 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 .
TOTAL_COST_SUB_ PROJECT	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.
PI_DEPARTMENT	Department of the PI
PI_DEPT_CODE	Code of the department
Flag	R represents RePorter's data, C represents Crisp's data, RC represents the result of integrating the two data sources in the years
PI_FIRST_NAME	The first name of PI
PI_MIDDLE_NAME	The middle name of PI
PI_LAST_NAME	The last name of PI
Year	The release year of the data
ORG_IPF_CODE	The Institution Profile (IPF) number is an internal NIH identifier that uniquely identifies and associates institutional information within NIH electronic systems. The NIH assigns an IPF number after the institution submits its request for registration.

17 B03_map_P MID_ProjID

The related information of the funded project and the funded article PMID published on the NIH official website.

Column Name	Description
NIH_P MID	Unique ID assigned by PubMed to identify PubMed articles.
PROJECT_NUMBER	An identifier for each research project, used to associate the project with publication and patent records. This identifier is not specific to any particular year of the project. It consists of

	the project activity code, administering IC, and serial number (a concatenation of Activity, Administering_IC, and Serial_Number).
PubYear	Publication year of the current article.
RecordYear	Year recorded in the NIH data file

18 B07_ORCID_Main

PubMed related author information extracted from ORCID

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Au_Order	Author order of the current author in the author list of current articles.
LastName	The last name of current author
ForeName	The name excluding the last name of current author
ORCID	Unique researcher ID that distinguishes the researcher from others allocated by ORCID.
AID	Unique author ID (The final author disambiguation result. S2id as the main source, supplemented by other disambiguation results)

19 B08_ORCID_Employment

Employment information of scientific personnel from ORCID dataset

Column Name	Description
ORCID	Unique researcher ID that distinguishes the researcher from others allocated by ORCID.
Department	The department which the researcher belongs to.
BeginYear	The beginning year of the researcher's employment.
Organization	The institution which the researcher belongs to.
City	The city where the researcher works.
Region	The region where the researcher works.
Country	The country where the researcher works.
Identifier	The identifier of an organization.
IdSource	The provider of an organizations' identifier.
EndYear	The end year of the researcher's employment.
AID	Unique author ID (The final author disambiguation result. S2id as the main source, supplemented by other disambiguation results)

20 B09_ORCID_Education

Education information of scientific personnel from ORCID dataset

Column Name	Description
ORCID	Unique researcher ID that distinguishes the researcher from others allocated by ORCID.
BeginYear	The beginning year of the researcher's education.
Organization	The organization the researcher has been educated.
City	The city that the author belongs to.
Region	The region that the author belongs to.
Country	The country that the author belongs to.
Identifier	The identifier of an organization.
IdSource	The provider of an organizations' identifier.
EndYear	The end year of the researcher's education.
Role	The degree that the researcher received.

21 B10_BERN_MAIN

Entity information set extracted from document titles and abstracts using BioBert

Column Name	Description
PMID	Unique ID assigned by PubMed to identify PubMed articles.
StartPosition	Start position of mention in an abstract.
EndPosition	End position of mention in an abstract.
Mention	Entity mentioned in an abstract.
EntityID	Normalized entity ID.
Type	Enumerated type of entity; values include species, disease, gene, drug, and mutation.

22 B11_BERN_EntityType

Entity Type dictionary

Column Name	Description
Id	Unique ID assigned to identify entity types, Corresponding to the field Type in table B10_BERN_Main..
Type	Entity type name.

23 B12_BERN_Mutation

Mutation information set extracted from literature titles and abstracts using BioBert

Column Name	Description
Main_id	Foreign key references from Bio-entities_Main (id).
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Mention	Mutation entity mentioned in the abstract.
MutationType	Normalized entity ID.
NormalizedName	Enumerated type of entity; values include species, disease, gene, drug, and mutation.

24 B14_Scimgo

Data Set of Bibliometric Indicators for Ranking Journals Based on Citation Source Information

Column Name	Description
Sourceid	Journal number provided by the dataset
Title	Title of the journal
Type	Type of journal or other publications
eISSN	Electronic ISSN
pISSN	Print ISSN number (may be the same as eISSN)
SJR	It refers to SCImago Journal Rankings It expresses the average number of weighted citations received in the selected year by the documents published in the selected journal in the three previous years, --i.e. weighted citations received in year X to documents published in the journal in years X-1, X-2 and X-3.
SJR_Best_Quartile	SJR partition
Hindex	The h index expresses the journal's number of articles (<i>h</i>) that have received at least <i>h</i> citations. It quantifies both journal scientific productivity and scientific impact and it is also applicable to scientists, countries, etc.
TotalDocs_toyear	Output of the selected year. All types of documents are considered, including citable and non citable documents.
TotalDocs_3years	Published documents in the three previous years (selected year documents are excluded), i.e., when the year X is selected, then X-1, X-2 and X-3 published documents are retrieved. All types of documents are considered, including citable and non-citable documents.
TotalRefs	It includes all the bibliographical references in a journal in the selected period.
TotalCities_3years	Number of citations received in the selected year by a journal to the documents published in the three previous years, --i.e. citations received in year X to documents published in years X-1, X-2 and X-

	3. All types of documents are considered.
CitableDocs_3years	Average citations per document in a 3 year period. It is computed considering the number of citations received by a journal in the current year to the documents published in the three previous years, -i.e. citations received in year X to documents published in years X-1, X-2 and X-3.
CitesPerDocs_2years	Average citations per document in a 2 year period. It is computed considering the number of citations received by a journal in the current year to the documents published in the two previous years, --i.e. citations received in year X to documents published in years X-1 and X-2.
RefPerDoc	Average number of references per document in the selected year.
Country	The country which the journal belongs to.
Publisher	The Publisher of the journal
Coverage	Year range of the journal
Categories	Subject Scope and SJR Division of the journal
ToYear	Publication year of the journal

25 C03_Affiliation_merge

Merging B05_Vetle_Map and A13_AffiliationList information, including the parsed organization information, such as Zip code, Location, Country, etc.

Column Name	Description
Id	Unique ID assigned to each record.
PMID	Unique ID assigned by PubMed to identify PubMed articles.
Au_Order	Author order of an article
AffiliationOrder	Affiliation order of one author, in case one author has multiple affiliations.
Affiliation	Affiliation text.
Department	Department parsed from Affiliation.
Institution	Institution parsed from Affiliation.
Email	Email parsed from Affiliation.
Zipcode	Zipcode parsed from Affiliation.
Location	Location parsed from Affiliation.
Country	Country parsed from Affiliation.
AID	Author disambiguation ID.
City	City parsed from Affiliation.
State	State parsed from Affiliation.
Vetle_Country	Country abbreviation.
Type	Affiliation type, such as COM, EDU, ORG, et.al.
Lat	Latitude of the Affiliation.

Lon	Longitude of the Affiliation.
Fips	Fips code.

26 C04_ReferenceList

The C04_ReferenceList contains 633401975 citations from 23856949 articles. The sources of data integration include PubMed's own citation data, NIH's opencitation collection, opencitations(run by David Shotton and Silvio Peroni) and the citation data from WOS. Compared with PubMed's own citation data (the amount of data is 223261597), it increased by 410140378. Compared with the previous version of PKG, the WOS citation data (the amount of data is 447596685), it increased by 185805290.

Column Name	Description
Id	Unique ID assigned to each record.
PMID	Unique ID assigned by PubMed to identify PubMed articles.
RefArticleId	PMIDs of references.

27 C05_NIH_PubMed

According to PMID and the author's name (full name of the last name and initials) from A02 and B02, match the author AID and NIH Principal Investigator (PI) number PIID to generate correspondence table C05, including PIID, AID, CORE_PROJECT_NUMBER, PIID, etc.

Column Name	Description
Id	Unique ID assigned to each record.
PI_IDS	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.
PMID	Unique ID assigned by PubMed to identify PubMed articles.
ProjectNumber	Project number.
SubProjectNumber	Sub project number.
AID	Author disambiguation ID.
PI_Name	The full name of PI.