

Google Patents Public Datasets

patents-public-data:patents.publications

Google Patents Public Data, provided by IFI CLAIMS Patent Services, is a worldwide bibliographic and US full-text dataset of patent publications.

“Google Patents Public Data” by IFI CLAIMS Patent Services and Google is licensed under a Creative Commons Attribution 4.0 International License.

Stat	Value
Last updated	2018-11-26
Rows	98,176,830
Size	899.4 GB

Schema

View in BigQuery

- **publication_number** STRING NULLABLE joins on **publication_number**
Patent publication number (DOCDB compatible), eg: ‘US-7650331-B1’
- **application_number** STRING NULLABLE joins on **application_number**
Patent application number (DOCDB compatible), eg: ‘US-87124404-A’. This may not always be set.
- **country_code** STRING NULLABLE
Country code, eg: ‘US’, ‘EP’, etc
- **kind_code** STRING NULLABLE
Kind code, indicating application, grant, search report, correction, etc. These are different for each country.
- **application_kind** STRING NULLABLE
High-level kind of the application: A=patent; U=utility; P=provision; W= PCT; F=design; T=translation.
- **application_number_formatted** STRING NULLABLE
Application number, formatted to the patent office format where possible.
- **pct_number** STRING NULLABLE
PCT number for this application if it was part of a PCT filing, eg: ‘PCT/EP2008/062623’.
- **family_id** STRING NULLABLE joins on **family_id**
Family ID (simple family). Grouping on family ID will return all publications associated with a simple patent family (all publications share the same priority claims).
- **title_localized** RECORD REPEATED
The publication titles in different languages
- **title_localized.text** STRING NULLABLE
Localized text
- **title_localized.language** STRING NULLABLE
Two-letter language code for this text
- **abstract_localized** RECORD REPEATED

The publication abstracts in different languages

- `abstract_localized.text` STRING NULLABLE
Localized text
- `abstract_localized.language` STRING NULLABLE
Two-letter language code for this text
- `claims_localized` RECORD REPEATED
For US publications only, the claims
- `claims_localized.text` STRING NULLABLE
Localized text
- `claims_localized.language` STRING NULLABLE
Two-letter language code for this text
- `description_localized` RECORD REPEATED
For US publications only, the description, limited to the first 9 megabytes
- `description_localized.text` STRING NULLABLE
Localized text
- `description_localized.language` STRING NULLABLE
Two-letter language code for this text
- `publication_date` INTEGER NULLABLE
The publication date.
- `filing_date` INTEGER NULLABLE
The filing date.
- `grant_date` INTEGER NULLABLE
The grant date, or 0 if not granted.
- `priority_date` INTEGER NULLABLE
The earliest priority date from the priority claims, or the filing date.
- `priority_claim` RECORD REPEATED
The application numbers of the priority claims of this publication.
- `priority_claim.publication_number` STRING NULLABLE
Same as `publication_number`
- `priority_claim.application_number` STRING NULLABLE
Same as `application_number`
- `priority_claim.npl_text` STRING NULLABLE
Free-text citation (non-patent literature, etc).
- `priority_claim.type` STRING NULLABLE
The type of reference (see parent field for values).
- `priority_claim.category` STRING NULLABLE
The category of reference (see parent field for values).
- `priority_claim.filing_date` INTEGER NULLABLE
The filing date.

- `inventor` STRING REPEATED
The inventors.
- `inventor_harmonized` RECORD REPEATED
The harmonized inventors and their countries.
- `inventor_harmonized.name` STRING NULLABLE
Name
- `inventor_harmonized.country_code` STRING NULLABLE
The two-letter country code
- `assignee` STRING REPEATED
The assignees/applicants.
- `assignee_harmonized` RECORD REPEATED
The harmonized assignees and their countries.
- `assignee_harmonized.name` STRING NULLABLE
Name
- `assignee_harmonized.country_code` STRING NULLABLE
The two-letter country code
- `examiner` RECORD REPEATED
The examiner of this publication and their countries.
- `examiner.name` STRING NULLABLE
Name
- `examiner.department` STRING NULLABLE
The examiner's department
- `examiner.level` STRING NULLABLE
The examiner's level
- `uspc` RECORD REPEATED
The US Patent Classification (USPC) codes.
- `uspc.code` STRING NULLABLE
Classification code
- `uspc.inventive` BOOLEAN NULLABLE
Is this classification inventive/main?
- `uspc.first` BOOLEAN NULLABLE
Is this classification the first/primary?
- `uspc.tree` STRING REPEATED
The full classification tree from the root to this code
- `ipc` RECORD REPEATED
The International Patent Classification (IPC) codes.
- `ipc.code` STRING NULLABLE
Classification code
- `ipc.inventive` BOOLEAN NULLABLE

- Is this classification inventive/main?
- **ipc.first** BOOLEAN NULLABLE
 - Is this classification the first/primary?
- **ipc.tree** STRING REPEATED
 - The full classification tree from the root to this code
- **cpc** RECORD REPEATED
 - The Cooperative Patent Classification (CPC) codes.
- **cpc.code** STRING NULLABLE
 - Classification code
- **cpc.inventive** BOOLEAN NULLABLE
 - Is this classification inventive/main?
- **cpc.first** BOOLEAN NULLABLE
 - Is this classification the first/primary?
- **cpc.tree** STRING REPEATED
 - The full classification tree from the root to this code
- **fi** RECORD REPEATED
 - The FI classification codes.
- **fi.code** STRING NULLABLE
 - Classification code
- **fi.inventive** BOOLEAN NULLABLE
 - Is this classification inventive/main?
- **fi.first** BOOLEAN NULLABLE
 - Is this classification the first/primary?
- **fi.tree** STRING REPEATED
 - The full classification tree from the root to this code
- **fterm** RECORD REPEATED
 - The F-term classification codes.
- **fterm.code** STRING NULLABLE
 - Classification code
- **fterm.inventive** BOOLEAN NULLABLE
 - Is this classification inventive/main?
- **fterm.first** BOOLEAN NULLABLE
 - Is this classification the first/primary?
- **fterm.tree** STRING REPEATED
 - The full classification tree from the root to this code
- **citation** RECORD REPEATED
 - The citations of this publication. Category is one of {CH2 = Chapter 2; SUP = Supplementary search report ; ISR = International search report ; SEA = Search report; APP = Applicant; EXA = Examiner; OPP = Opposition; 115 = article 115; PRS = Pre-grant pre-search; APL = Appealed; FOP = Filed opposition}, Type is one of {A = technological background; D = document cited in application; E = earlier patent document; 1 = document cited for other reasons; O = Non-written disclosure; P = Intermediate

document; T = theory or principle; X = relevant if taken alone; Y = relevant if combined with other documents}

- `citation.publication_number` STRING NULLABLE
Same as `publication_number`
- `citation.application_number` STRING NULLABLE
Same as `application_number`
- `citation.npl_text` STRING NULLABLE
Free-text citation (non-patent literature, etc).
- `citation.type` STRING NULLABLE
The type of reference (see parent field for values).
- `citation.category` STRING NULLABLE
The category of reference (see parent field for values).
- `citation.filing_date` INTEGER NULLABLE
The filing date.
- `entity_status` STRING NULLABLE
The USPTO entity status (large, small).
- `art_unit` STRING NULLABLE
The USPTO art unit performing the examination (2159, etc).

Join columns

`publication_number`

joins to `patents-public-data:google_patents_research.publications::publication_number` on `publication_number` (100.00%, 98,176,830 rows)

Key	Percent	Rows	Sample values
CA	100.00%	3,178,287	['CA-2802720-A1', 'CA-2945439-A1', 'CA-1103736-A', 'CA-1316706-C', 'CA-935404-A']
FR	100.00%	3,148,849	['FR-2335581-B1', 'FR-3042405-B3', 'FR-2377701-B1', 'FR-673660-A', 'FR-3025880-A1']
RU	100.00%	1,260,195	['RU-2002118251-A', 'RU-2012153197-A', 'RU-2006145355-A', 'RU-1167918-C', 'RU-2495773-C2']
WO	100.00%	4,053,857	['WO-2012091842-A3', 'WO-2014097019-A1', 'WO-2008039352-A3', 'WO-2008119526-A1', 'WO-2009014098-A1']
US	100.00%	16,086,541	['US-7135945-B2', 'US-2014229180-A1', 'US-2006027353-A1', 'US-5247217-A', 'US-3519113-A']
BE	100.00%	791,842	['BE-604027-A1', 'BE-783201-A1', 'BE-581533-A', 'BE-412802-A', 'BE-904680-A']
DE	100.00%	7,620,394	['DE-3227556-C1', 'DE-3010674-A1', 'DE-2401062-C2', 'DE-2147964-A1', 'DE-20117122-U1']
NL	100.00%	644,372	['NL-169751-B', 'NL-7314201-A', 'NL-1024649-C1', 'NL-6700965-A', 'NL-285217-A']
ES	100.00%	1,648,727	['ES-8604715-A1', 'ES-2644791-T3', 'ES-2239165-T3', 'ES-2543589-T3', 'ES-1007126-U']
GB	100.00%	3,721,376	['GB-2108723-B', 'GB-201502035-D0', 'GB-2495100-A', 'GB-2482468-B', 'GB-275149-A']
DK	100.00%	579,908	['DK-90747-C', 'DK-58287-C', 'DK-54275-C', 'DK-2446511-T3', 'DK-98061-C']

Key	Percent	Rows	Sample values
KR	100.00%	4,810,158	['KR-20110045650-A', 'KR-20030087108-A', 'KR-101712592-B1', 'KR-20110011402-U', 'KR-20070090061-A']
EP	100.00%	6,294,864	['EP-1767747-A3', 'EP-2409007-A4', 'EP-0347381-A1', 'EP-0596467-A1', 'EP-2670850-A1']
JP	100.00%	24,412,531	['JP-S617693-B2', 'JP-2015088527-A', 'JP-S56149143-A', 'JP-S5065977-U', 'JP-2003125918-A']
FI	100.00%	617,048	['FI-104435-B1', 'FI-890200-D0', 'FI-20095967-D0', 'FI-823726-A0', 'FI-105615-B']
CN	100.00%	19,229,782	['CN-1636837-A', 'CN-202884563-U', 'CN-101058397-A', 'CN-101800285-A', 'CN-207377654-U']
LU	100.00%	78,099	['LU-72898-A1', 'LU-53593-A1', 'LU-54195-A1', 'LU-54084-A1', 'LU-42030-A1']

#standardSQL

```
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.google_patents_research.publications`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3
```

joins to patents-public-data:patentsview.match::publication_number on **publication_number** (6.47%, 6,353,172 rows)

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-2864208-A1', 'CA-993509-A', 'CA-447265-A', 'CA-2049617-A1', 'CA-2223309-A1']
FR	<i>none</i>	0	['FR-874553-A', 'FR-2558087-B1', 'FR-2994149-B1', 'FR-2530634-B1', 'FR-2269315-B1']
RU	<i>none</i>	0	['RU-2323533-C2', 'RU-2276315-C2', 'RU-2008106365-A', 'RU-2157139-C2', 'RU-2417081-C1']
WO	<i>none</i>	0	['WO-2004090020-A1', 'WO-2017023940-A1', 'WO-2017094525-A1', 'WO-2016054232-A1', 'WO-2011092083-A3']
US	39.49%	6,353,172	['US-4902986-A', 'US-2182405-A', 'US-9223134-B2', 'US-2015045005-A1', 'US-2566296-A']
BE	<i>none</i>	0	['BE-857959-A1', 'BE-424746-A', 'BE-608278-A', 'BE-873779-A', 'BE-872277-A1']
DE	<i>none</i>	0	['DE-1126821-B', 'DE-2310537-A1', 'DE-3120101-A1', 'DE-2365499-A1', 'DE-102011053898-A1']
NL	<i>none</i>	0	['NL-194728-B', 'NL-109015-C', 'NL-7904995-A', 'NL-7714516-A', 'NL-293657-A']
ES	<i>none</i>	0	['ES-488354-D0', 'ES-295473-U', 'ES-214203-U', 'ES-262909-A1', 'ES-166123-Y']
LU	<i>none</i>	0	['LU-41343-A1', 'LU-52331-A1', 'LU-75696-A1', 'LU-77684-A1', 'LU-84871-A1']
DK	<i>none</i>	0	['DK-119687-D0', 'DK-0888481-T3', 'DK-137712-B', 'DK-647950-T3', 'DK-16608-C']
KR	<i>none</i>	0	['KR-101721381-B1', 'KR-20110126522-A', 'KR-20110049768-A', 'KR-20000069614-A', 'KR-20070096871-A']
EP	<i>none</i>	0	['EP-0265549-B1', 'EP-2201006-A1', 'EP-0744082-A1', 'EP-0961442-B1', 'EP-3245178-A1']

Key	Percent	Rows	Sample values
JP	<i>none</i>	0	['JP-2000025454-A', 'JP-2001194532-A', 'JP-2006308848-A', 'JP-H0635832-U', 'JP-S5086696-A']
GB	<i>none</i>	0	['GB-739779-A', 'GB-1121875-A', 'GB-2512043-A', 'GB-1467440-A', 'GB-9723293-D0']
FI	<i>none</i>	0	['FI-854801-A0', 'FI-32880-A', 'FI-75901-B', 'FI-31533-A', 'FI-896320-A0']
CN	<i>none</i>	0	['CN-207595440-U', 'CN-103406813-B', 'CN-103599329-A', 'CN-207303843-U', 'CN-101590077-B']

#standardSQL

SELECT

```

COUNT(*) AS cnt,
COUNT(second.second_column) AS second_cnt,
first.country_code AS grouped,
ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value

```

FROM `patents-public-data.patents.publications` AS first

LEFT JOIN (

```

SELECT publication_number AS second_column, COUNT(*) AS cnt
FROM `patents-public-data.patentsview.match`
GROUP BY 1

```

) AS second ON first.publication_number = second.second_column

GROUP BY 3

joins to patents-public-data:uspto_oce_assignment.match::publication_number on publication_number (9.03%, 8,869,739 rows)

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-749155-A', 'CA-2079720-A1', 'CA-2663677-C', 'CA-2083320-A1', 'CA-485145-A']
LU	<i>none</i>	0	['LU-38866-A1', 'LU-29584-A1', 'LU-75688-A1', 'LU-40043-A', 'LU-86865-A1']
RU	<i>none</i>	0	['RU-2007128299-A', 'RU-2011101500-A', 'RU-2016590-C1', 'RU-2097763-C1', 'RU-2410772-C1']
WO	<i>none</i>	0	['WO-2005120908-A1', 'WO-9416915-A1', 'WO-2016020424-A1', 'WO-2016190210-A1', 'WO-2010037138-A3']
US	55.14%	8,869,739	['US-3237743-A', 'US-2018179808-A1', 'US-7517268-B2', 'US-2013118072-A1', 'US-2318472-A']
BE	<i>none</i>	0	['BE-842529-A', 'BE-449847-A', 'BE-613692-A1', 'BE-809459-A1', 'BE-844135-A1']
DE	<i>none</i>	0	['DE-1360231-U', 'DE-2843149-A1', 'DE-4009999-A1', 'DE-1447834-A1', 'DE-102016105670-A1']
EP	<i>none</i>	0	['EP-3083443-B1', 'EP-0907515-B1', 'EP-3180819-A4', 'EP-2484707-A1', 'EP-0914350-A4']
ES	<i>none</i>	0	['ES-2043016-T3', 'ES-490320-D0', 'ES-1002631-U', 'ES-2010751-A6', 'ES-216314-Y']
GB	<i>none</i>	0	['GB-0126912-D0', 'GB-1093592-A', 'GB-190512133-A', 'GB-1317741-A', 'GB-203777-A']
KR	<i>none</i>	0	['KR-0116122-Y1', 'KR-20180019685-A', 'KR-20000040131-A', 'KR-20060017552-A', 'KR-20170064796-A']
DK	<i>none</i>	0	['DK-0801901-T3', 'DK-601469-T3', 'DK-0889885-T3', 'DK-108908-C', 'DK-544086-A']
JP	<i>none</i>	0	['JP-2018020213-A', 'JP-H0446577-B2', 'JP-3438636-B2', 'JP-2006272987-A', 'JP-2012020547-A']
FI	<i>none</i>	0	['FI-841820-A0', 'FI-78711-B', 'FI-109605-B1', 'FI-20070229-D0', 'FI-930459-V0']
FR	<i>none</i>	0	['FR-1238614-A', 'FR-478802-A', 'FR-2537934-A1', 'FR-2232543-A1', 'FR-2186285-A1']

Key	Percent	Rows	Sample values
NL	<i>none</i>	0	['NL-108801-C', 'NL-6801358-A', 'NL-6508430-A', 'NL-7608071-A', 'NL-20764-C']
CN	<i>none</i>	0	['CN-2488552-Y', 'CN-105455417-A', 'CN-1057111-A', 'CN-205089197-U', 'CN-203635647-U']

#standardSQL

SELECT

```

COUNT(*) AS cnt,
COUNT(second.second_column) AS second_cnt,
first.country_code AS grouped,
ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first

```

LEFT JOIN (

```

SELECT publication_number AS second_column, COUNT(*) AS cnt
FROM `patents-public-data.uspto_oce_assignment.match`
GROUP BY 1

```

) AS second ON first.publication_number = second.second_column

GROUP BY 3

joins to patents-public-data:uspto_oce_cancer.match::publication_number on **publication_number** (0.27%, 269,168 rows)

Key	Percent	Rows	Sample values
GB	<i>none</i>	0	['GB-1189382-A', 'GB-991033-A', 'GB-9418292-D0', 'GB-9503073-D0', 'GB-2296867-A']
FR	<i>none</i>	0	['FR-423116-A', 'FR-2859211-A1', 'FR-464248-A', 'FR-2058161-A5', 'FR-29252-E']
RU	<i>none</i>	0	['RU-2140676-C1', 'RU-2396997-C1', 'RU-2660616-C2', 'RU-2527333-C1', 'RU-2264251-C2']
WO	<i>none</i>	0	['WO-9946592-A1', 'WO-2006065885-A3', 'WO-0143571-A1', 'WO-2014106570-A1', 'WO-0111296-A2']
US	1.67%	269,168	['US-2014348367-A1', 'US-2013193224-A1', 'US-2016337213-A1', 'US-2008306656-A1', 'US-1200636-A']
BE	<i>none</i>	0	['BE-757866-A', 'BE-502458-A', 'BE-466627-A', 'BE-599998-A', 'BE-767254-A1']
DE	<i>none</i>	0	['DE-219404-C', 'DE-882149-C', 'DE-4316776-A1', 'DE-3039689-C2', 'DE-102009053732-A1']
NL	<i>none</i>	0	['NL-8500672-A', 'NL-2003093-C2', 'NL-7600598-A', 'NL-171156-C', 'NL-73637-C']
ES	<i>none</i>	0	['ES-1013045-U', 'ES-156216-U', 'ES-2046753-T3', 'ES-2246235-T3', 'ES-2049165-A1']
LU	<i>none</i>	0	['LU-47660-A', 'LU-30765-A1', 'LU-45511-A', 'LU-46897-A1', 'LU-35139-A']
DK	<i>none</i>	0	['DK-132852-A', 'DK-137492-B', 'DK-0550636-T3', 'DK-142880-B', 'DK-600378-T3']
KR	<i>none</i>	0	['KR-20080103231-A', 'KR-20150086579-A', 'KR-100405569-B1', 'KR-20140006049-U', 'KR-100564637-B1']
EP	<i>none</i>	0	['EP-2140623-B1', 'EP-1604623-B1', 'EP-1092717-B1', 'EP-2720733-A2', 'EP-1793674-A4']
JP	<i>none</i>	0	['JP-2008287862-A', 'JP-2003510307-A', 'JP-2004230446-A', 'JP-H09112357-A', 'JP-4108517-B2']
FI	<i>none</i>	0	['FI-915424-A', 'FI-862887-A', 'FI-803574-L', 'FI-844536-A0', 'FI-781009-A']
CN	<i>none</i>	0	['CN-105328743-A', 'CN-101492845-A', 'CN-108390314-A', 'CN-103020698-A', 'CN-207323046-U']
CA	<i>none</i>	0	['CA-1085195-A', 'CA-543244-A', 'CA-564775-A', 'CA-325095-A', 'CA-730138-A']

```
#standardSQL
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.uspto_oce_cancer.match`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

joins to patents-public-data:uspto_oce_claims.match::publication_number on publication_number (9.12%,
8,950,539 rows)
```

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-645177-A', 'CA-1117656-A', 'CA-1020324-A', 'CA-2118551-A1', 'CA-2643178-C']
LU	<i>none</i>	0	['LU-57673-A1', 'LU-64045-A1', 'LU-72131-A1', 'LU-66575-A1', 'LU-80367-A1']
RU	<i>none</i>	0	['RU-2018112718-A', 'RU-2601924-C2', 'RU-2003102240-A', 'RU-2396065-C2', 'RU-2004135245-A']
WO	<i>none</i>	0	['WO-2007030966-A2', 'WO-9829454-A1', 'WO-9611958-A1', 'WO-0216104-A3', 'WO-2013052427-A2']
US	55.64%	8,950,539	['US-911753-A', 'US-9247740-B2', 'US-2013315081-A1', 'US-1150182-A', 'US-2017063410-A1']
BE	<i>none</i>	0	['BE-900478-A1', 'BE-814101-A', 'BE-716191-A', 'BE-771621-A1', 'BE-439674-A']
DE	<i>none</i>	0	['DE-2914465-C2', 'DE-102007028729-B4', 'DE-69220872-D1', 'DE-1924191-U', 'DE-817802-C']
NL	<i>none</i>	0	['NL-6509782-A', 'NL-292818-A', 'NL-12656-C', 'NL-6916681-A', 'NL-40640-C']
ES	<i>none</i>	0	['ES-2552460-T3', 'ES-2074011-A2', 'ES-295141-A1', 'ES-8701689-A1', 'ES-502815-D0']
EP	<i>none</i>	0	['EP-0594480-B1', 'EP-0711384-A1', 'EP-0974212-B1', 'EP-3010711-A1', 'EP-0155950-A4']
GB	<i>none</i>	0	['GB-2513311-A', 'GB-112934-A', 'GB-472166-A', 'GB-2431025-B', 'GB-1383893-A']
KR	<i>none</i>	0	['KR-100399752-B1', 'KR-20170114300-A', 'KR-20110054710-A', 'KR-950032887-U', 'KR-20140000587-A']
DK	<i>none</i>	0	['DK-67357-C', 'DK-1080054-T3', 'DK-601686-D0', 'DK-93140-C', 'DK-20403-C']
JP	<i>none</i>	0	['JP-H03282757-A', 'JP-2008281068-A', 'JP-S5516398-B2', 'JP-H0533535-A', 'JP-2009194162-A']
FI	<i>none</i>	0	['FI-20155329-A', 'FI-86160-B', 'FI-102386-B', 'FI-U920066-U0', 'FI-905025-D0']
FR	<i>none</i>	0	['FR-346608-A', 'FR-2710691-A1', 'FR-2202175-A1', 'FR-1530962-A', 'FR-771138-A']
CN	<i>none</i>	0	['CN-108053489-A', 'CN-107360646-A', 'CN-106852430-A', 'CN-204360485-U', 'CN-102378650-A']

```
#standardSQL
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
```

```

    ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
    SELECT publication_number AS second_column, COUNT(*) AS cnt
    FROM `patents-public-data.uspto_oce_claims.match`
    GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

joins to patents-public-data:usitc_investigations.match::publication_number on publication_number
(0.00%, 1,419 rows)

```

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-2496125-A1', 'CA-2701416-A1', 'CA-611882-A', 'CA-449278-A', 'CA-561220-A']
BE	<i>none</i>	0	['BE-837675-A1', 'BE-532052-A', 'BE-801672-A1', 'BE-844330-A1', 'BE-883734-A']
FR	<i>none</i>	0	['FR-2075925-A1', 'FR-2274765-B1', 'FR-841839-A', 'FR-849715-A', 'FR-934554-A']
RU	<i>none</i>	0	['RU-2038691-C1', 'RU-2239328-C1', 'RU-2005135231-A', 'RU-94043324-A', 'RU-2606395-C2']
WO	<i>none</i>	0	['WO-2009025580-A1', 'WO-2012129262-A1', 'WO-2017217126-A1', 'WO-9964619-A3', 'WO-9703931-A1']
US	0.01%	1,419	['US-2011156832-A1', 'US-790870-A', 'US-2006198508-A1', 'US-2017079750-A1', 'US-D456723-S']
ES	<i>none</i>	0	['ES-2021078-B3', 'ES-2192427-A1', 'ES-2160086-A1', 'ES-529178-A0', 'ES-527050-D0']
GB	<i>none</i>	0	['GB-1181687-A', 'GB-2161282-A', 'GB-2368674-A', 'GB-2389683-B', 'GB-2253281-B']
EP	<i>none</i>	0	['EP-2248870-A3', 'EP-1295264-B1', 'EP-2757397-A1', 'EP-2678356-B1', 'EP-0221863-A3']
DE	<i>none</i>	0	['DE-1939451-U', 'DE-7536813-U', 'DE-1674797-U', 'DE-3666489-D1', 'DE-3042053-C1']
LU	<i>none</i>	0	['LU-57899-A1', 'LU-50413-A1', 'LU-62786-A1', 'LU-49559-A', 'LU-78371-A1']
KR	<i>none</i>	0	['KR-950007812-Y1', 'KR-200433231-Y1', 'KR-200232050-Y1', 'KR-20080042224-A', 'KR-100423713-B1']
DK	<i>none</i>	0	['DK-136056-B', 'DK-50542-C', 'DK-143193-D0', 'DK-172476-A', 'DK-287690-D0']
JP	<i>none</i>	0	['JP-S5475913-U', 'JP-2939106-B2', 'JP-S57157091-U', 'JP-H0246976-Y2', 'JP-S6352647-B2']
FI	<i>none</i>	0	['FI-971726-D0', 'FI-901626-A0', 'FI-66613-B', 'FI-82421-C', 'FI-47843-B']
NL	<i>none</i>	0	['NL-6803640-A', 'NL-270012-A', 'NL-34351-C', 'NL-2006378-C', 'NL-167204-B']
CN	<i>none</i>	0	['CN-1257949-C', 'CN-206904078-U', 'CN-205584795-U', 'CN-107267188-A', 'CN-103180930-A']

```

#standardSQL
SELECT
    COUNT(*) AS cnt,
    COUNT(second.second_column) AS second_cnt,
    first.country_code AS grouped,
    ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
    SELECT publication_number AS second_column, COUNT(*) AS cnt
    FROM `patents-public-data.usitc_investigations.match`

```

```

GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3
joins to patents-public-data:marec.publications::publication_number on publication_number (13.83%,
13,577,285 rows)

```

Key	Percent	Rows	Sample values
CA	none	0	['CA-1080903-A1', 'CA-2400431-A1', 'CA-1118501-A1', 'CA-2596999-A1', 'CA-2879356-A1']
FR	none	0	['FR-978523-A', 'FR-1056540-A', 'FR-2312743-A1', 'FR-2436663-B1', 'FR-399372-A']
RU	none	0	['RU-2371294-C2', 'RU-37567-U1', 'RU-2437708-C1', 'RU-2349278-C1', 'RU-2447125-C1']
WO	44.03%	1,784,957	['WO-0221414-A1', 'WO-2013127963-A1', 'WO-0215126-A3', 'WO-2012040115-A1', 'WO-2008058906-A3']
US	34.85%	5,606,641	['US-2305030-A', 'US-3110812-A', 'US-2003191655-A1', 'US-1856688-A', 'US-2009094344-A1']
BE	none	0	['BE-449234-A', 'BE-656674-A', 'BE-742009-A', 'BE-822538-A1', 'BE-895139-A1']
DE	none	0	['DE-69624663-T2', 'DE-19801321-A1', 'DE-1342480-U', 'DE-412965-C', 'DE-2813452-A1']
DK	none	0	['DK-158639-B', 'DK-0388596-T3', 'DK-45136-C', 'DK-36457-C', 'DK-0704438-T3']
ES	none	0	['ES-2014428-B3', 'ES-2571431-T3', 'ES-453400-A1', 'ES-184793-U', 'ES-522456-D0']
EP	55.72%	3,507,653	['EP-2402742-B1', 'EP-0442768-A3', 'EP-2895855-A1', 'EP-0685438-A1', 'EP-1614942-B1']
LU	none	0	['LU-31389-A1', 'LU-65787-A1', 'LU-58636-A1', 'LU-47487-A', 'LU-42685-A1']
KR	none	0	['KR-100760249-B1', 'KR-101183456-B1', 'KR-20060125244-A', 'KR-100904437-B1', 'KR-20120074168-A']
NL	none	0	['NL-6606121-A', 'NL-8304403-A', 'NL-6909057-A', 'NL-27634-C', 'NL-8103269-A']
JP	10.97%	2,678,034	['JP-H0222421-B2', 'JP-S5165361-A', 'JP-H08296637-A', 'JP-S5427761-Y2', 'JP-S5769549-U']
GB	none	0	['GB-2398927-B', 'GB-0220429-D0', 'GB-8304239-D0', 'GB-602131-A', 'GB-2022703-B']
FI	none	0	['FI-11162-U1', 'FI-71105-C', 'FI-2395-U1', 'FI-910448-D0', 'FI-962669-D0']
CN	none	0	['CN-103257828-A', 'CN-1631210-A', 'CN-202481858-U', 'CN-101777368-B', 'CN-103149057-A']

```

#standardSQL
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.marec.publications`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

```

```

joins to patents-public-data:ebi_surechembl.match::publication_number on publication_number (4.29%,
4,215,650 rows)

```

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-2353905-A1', 'CA-1124757-A', 'CA-1143705-A1', 'CA-2711943-C', 'CA-423703-A']
FR	<i>none</i>	0	['FR-2863560-B1', 'FR-2769749-B1', 'FR-2928603-A1', 'FR-1381072-A', 'FR-976320-A']
RU	<i>none</i>	0	['RU-2098533-C1', 'RU-16851-U1', 'RU-2251610-C2', 'RU-2018248-C1', 'RU-2439160-C1']
WO	19.64%	796,266	['WO-8800418-A1', 'WO-02087542-A1', 'WO-2004084576-A3', 'WO-02059635-A2', 'WO-2015005787-A1']
US	11.46%	1,843,383	['US-6962029-B2', 'US-9350343-B2', 'US-2016270536-A1', 'US-2017080945-A1', 'US-3760182-A']
BE	<i>none</i>	0	['BE-782007-A', 'BE-615483-R', 'BE-818542-A1', 'BE-802250-A1', 'BE-581938-A']
DE	<i>none</i>	0	['DE-2904658-C2', 'DE-102011119344-A1', 'DE-9111797-U1', 'DE-102011107539-A1', 'DE-34747-C']
NL	<i>none</i>	0	['NL-81865-C', 'NL-50186-C', 'NL-201166-A', 'NL-1002534-C1', 'NL-6616052-A']
ES	<i>none</i>	0	['ES-549957-D0', 'ES-220487-A1', 'ES-130979-Y', 'ES-2442391-A1', 'ES-1029542-U']
LU	<i>none</i>	0	['LU-51243-A1', 'LU-48096-A1', 'LU-48056-A1', 'LU-46758-A1', 'LU-27761-A1']
DK	<i>none</i>	0	['DK-1105096-T3', 'DK-1095688-T3', 'DK-0601130-T3', 'DK-60551-C', 'DK-112525-B']
KR	<i>none</i>	0	['KR-20130032191-A', 'KR-101000291-B1', 'KR-19980019516-U', 'KR-890000191-Y1', 'KR-900004490-B1']
EP	20.22%	1,272,625	['EP-3145888-A1', 'EP-1337330-A2', 'EP-0940901-B1', 'EP-2073121-A3', 'EP-1236120-A1']
JP	1.24%	303,376	['JP-H062767-Y2', 'JP-S563275-A', 'JP-S6117782-B2', 'JP-2005288458-A', 'JP-5505234-B2']
GB	<i>none</i>	0	['GB-2301685-A', 'GB-712922-A', 'GB-201720267-D0', 'GB-151510-A', 'GB-443778-A']
FI	<i>none</i>	0	['FI-944469-D0', 'FI-112756-B', 'FI-974222-D0', 'FI-964388-A0', 'FI-925796-A0']
CN	<i>none</i>	0	['CN-101356232-B', 'CN-203212527-U', 'CN-105527911-A', 'CN-207179339-U', 'CN-103777160-B']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.ebi_surechembl.match`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

```

joins to patents-public-data:ebi_chembl.match_24::publication_number on **publication_number** (0.00%, 3,963 rows)

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-883140-A', 'CA-758817-A', 'CA-2300020-C', 'CA-1281569-C', 'CA-2339278-A1']
FR	<i>none</i>	0	['FR-1270310-A', 'FR-539361-A', 'FR-1389340-A', 'FR-2966843-A1', 'FR-951537-A']

Key	Percent	Rows	Sample values
RU	<i>none</i>	0	['RU-2601967-C1', 'RU-2009115983-A', 'RU-2342607-C1', 'RU-95105969-A', 'RU-2014124465-A']
WO	<i>none</i>	0	['WO-2012096658-A1', 'WO-2016149302-A1', 'WO-2018072221-A1', 'WO-2014137564-A1', 'WO-2017039985-A1']
US	0.02%	3,963	['US-6250392-B1', 'US-2018210982-A1', 'US-4607609-A', 'US-2015364213-A1', 'US-2001054273-A1']
BE	<i>none</i>	0	['BE-404232-A', 'BE-811732-A1', 'BE-564076-A', 'BE-1006476-A3', 'BE-772646-A']
GB	<i>none</i>	0	['GB-9804087-D0', 'GB-634340-A', 'GB-9116526-D0', 'GB-2271535-B', 'GB-1447315-A']
EP	<i>none</i>	0	['EP-1138217-A1', 'EP-1672908-A1', 'EP-0692379-A3', 'EP-2824057-B1', 'EP-2310737-B1']
ES	<i>none</i>	0	['ES-45527-U', 'ES-2172743-T3', 'ES-2630203-T3', 'ES-2323278-T3', 'ES-1061574-U']
LU	<i>none</i>	0	['LU-80254-A1', 'LU-63420-A1', 'LU-57611-A1', 'LU-51881-A', 'LU-40507-A1']
KR	<i>none</i>	0	['KR-101399051-B1', 'KR-20050096041-A', 'KR-101423327-B1', 'KR-100640207-B1', 'KR-20140023306-A']
DK	<i>none</i>	0	['DK-1404844-T3', 'DK-138968-B', 'DK-2190443-T3', 'DK-154380-C', 'DK-135232-C']
JP	<i>none</i>	0	['JP-5725731-B2', 'JP-S5016470-B1', 'JP-4380275-B2', 'JP-2015118683-A', 'JP-S4940395-B1']
FI	<i>none</i>	0	['FI-87998-C', 'FI-4839-U1', 'FI-349-A', 'FI-823176-L', 'FI-20030855-A0']
NL	<i>none</i>	0	['NL-120017-A', 'NL-1005111-C2', 'NL-91623-C', 'NL-8101548-A', 'NL-7305975-A']
CN	<i>none</i>	0	['CN-101829429-B', 'CN-201444068-U', 'CN-107117262-A', 'CN-2360889-Y', 'CN-102643904-B']
DE	<i>none</i>	0	['DE-1784646-U', 'DE-693936-C', 'DE-2740809-A1', 'DE-20008386-U1', 'DE-2100034-A1']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.ebi_chembl.match_24`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

```

joins to innography-174118:technical_standards.etsi::PublicationNumber on **publication_number** (0.04%, 34,458 rows)

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-2219733-A1', 'CA-1288951-C', 'CA-947054-A', 'CA-2479985-A1', 'CA-1024991-A1']
FR	<i>none</i>	0	['FR-1223470-A', 'FR-2333172-A1', 'FR-564979-A', 'FR-2798597-A1', 'FR-1552969-A']
RU	<i>none</i>	0	['RU-2266701-C2', 'RU-2445629-C1', 'RU-2005140673-A', 'RU-2004108921-A', 'RU-2011129489-A']
WO	<i>none</i>	0	['WO-2012014953-A1', 'WO-2013190485-A1', 'WO-2009080010-A1', 'WO-2017028298-A1', 'WO-2010150552-A1']

Key	Percent	Rows	Sample values
US	0.21%	34,458	['US-8586724-B2', 'US-2009225022-A1', 'US-8535171-B2', 'US-581494-A', 'US-714937-A']
BE	<i>none</i>	0	['BE-902742-A', 'BE-761538-A1', 'BE-831255-A', 'BE-898179-A1', 'BE-337724-A']
LU	<i>none</i>	0	['LU-36991-A', 'LU-68599-A1', 'LU-72232-A1', 'LU-57105-A1', 'LU-50089-A1']
NL	<i>none</i>	0	['NL-2001377-C2', 'NL-1017636-C2', 'NL-288870-A', 'NL-1034119-C2', 'NL-44406-C']
ES	<i>none</i>	0	['ES-327146-A1', 'ES-254421-A1', 'ES-533817-D0', 'ES-201783-A2', 'ES-2253973-B1']
GB	<i>none</i>	0	['GB-189715-A', 'GB-966725-A', 'GB-322620-A', 'GB-479760-A', 'GB-2160021-A']
DK	<i>none</i>	0	['DK-104430-C', 'DK-2830682-T3', 'DK-0456350-T3', 'DK-288283-A', 'DK-129936-B']
KR	<i>none</i>	0	['KR-138312-Y1', 'KR-20140074214-A', 'KR-20060124583-A', 'KR-100639933-B1', 'KR-20000015897-A']
EP	<i>none</i>	0	['EP-0475696-A3', 'EP-2981275-A1', 'EP-2761267-B1', 'EP-0340787-A3', 'EP-2098640-A2']
JP	<i>none</i>	0	['JP-S56116254-U', 'JP-S59118312-U', 'JP-2009053189-A', 'JP-2004352722-A', 'JP-H09285342-A']
FI	<i>none</i>	0	['FI-107689-B', 'FI-100669-B', 'FI-3138-A', 'FI-822969-L', 'FI-48551-B']
CN	<i>none</i>	0	['CN-206784689-U', 'CN-202558328-U', 'CN-107809819-A', 'CN-106624467-A', 'CN-107356061-A']
DE	<i>none</i>	0	['DE-7702944-U1', 'DE-4446578-C2', 'DE-69432422-T2', 'DE-1885711-U', 'DE-10239447-B4']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT PublicationNumber AS second_column, COUNT(*) AS cnt
  FROM `innography-174118.technical_standards.etsi`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
GROUP BY 3

joins from patents-public-data:google_patents_research.publications::publication_number on publica-
tion_number (100.00%, 98,176,830 rows)

joins from patents-public-data:patentsview.match::publication_number on publication_number (99.93%,
6,353,198 rows)

joins from patents-public-data:uspto_oce_assignment.match::publication_number on publication_number
(99.96%, 8,869,754 rows)

joins from patents-public-data:uspto_oce_cancer.match::publication_number on publication_number (99.97%,
269,168 rows)

joins from patents-public-data:uspto_oce_claims.match::publication_number on publication_number (99.95%,
8,950,539 rows)

joins from patents-public-data:usitc_investigations.match::publication_number on publication_number
(99.93%, 1,419 rows)

joins from patents-public-data:marec.publications::publication_number on publication_number (71.12%,

```

13,585,005 rows)

joins from patents-public-data:ebi_surechembl.match::publication_number on publication_number (99.73%, 4,215,650 rows)

joins from patents-public-data:ebi_chembl.match_24::publication_number on publication_number (100.00%, 4,456 rows)

joins from innography-174118:technical_standards.etsi::PublicationNumber on publication_number (99.98%, 34,458 rows)

application_number

joins to patents-public-data:uspto_oce_office_actions.match_app::application_number on application_number (3.62%, 3,555,089 rows)

Key	Percent	Rows	Sample values
CA	none	0	['CA-2049312-A', 'CA-2611112-A', '', 'CA-185022-A', 'CA-221673-A']
FR	none	0	['FR-0652132-A', 'FR-904391D-A', 'FR-551898D-A', 'FR-974800D-A', 'FR-1451345-A']
RU	none	0	['RU-2007108428-A', 'RU-2011127626-U', 'RU-99126007-U', 'RU-2007123754-A', 'RU-2006108392-A']
WO	none	0	['US-2011023789-W', 'US-2006006578-W', 'US-9710335-W', 'US-2012050976-W', 'FR-2012053090-W']
US	22.10%	3,555,089	['US-96774307-A', 'US-201113017029-A', 'US-50184405-A', 'US-201113811216-A', 'US-1322501-A']
BE	none	0	['', 'BE-633072D-A', 'BE-544496D-A', 'BE-419615D-A', 'BE-382689D-A']
LU	none	0	['LU-71962-A', 'LU-35597D-A', 'LU-37602D-A', 'LU-74003-A', 'LU-34286D-A']
NL	none	0	['NL-37893D-A', 'NL-8004402-A', 'NL-259867D-A', 'NL-7404210-A', 'NL-253872-A']
ES	none	0	['ES-480386-A', 'ES-421053-A', 'ES-200600543-U', 'ES-187150-U', 'ES-99109986-T']
EP	none	0	['EP-00923525-A', 'EP-92400569-A', 'EP-91870122-A', 'EP-04252500-A', 'EP-15803369-A']
GB	none	0	['GB-9727254-A', 'GB-0415270-A', 'GB-201314499-A', 'GB-8529616-A', 'GB-1918652-A']
KR	none	0	['KR-20140192233-A', 'KR-20080103596-A', 'KR-20060103230-A', 'KR-20110044553-A', 'KR-20177011961-A']
DK	none	0	['DK-505984-A', 'DK-03778108-T', 'DK-225989-A', 'DK-576682-A', 'DK-09802199-T']
JP	none	0	['JP-10676982-U', 'JP-7487980-U', 'JP-26045791-A', 'JP-2009554399-A', 'JP-2000286745-A']
FI	none	0	['FI-U20050143-U', 'FI-951773-A', 'FI-932275-A', '', 'FI-20090382-A']
CN	none	0	['CN-201420032886-U', 'CN-201711083518-A', 'CN-201510121028-A', 'CN-201610012267-A', 'CN-201510935048-A']
DE	none	0	['DE-50011530-A', 'DE-2226015-A', 'DE-68926392-T', 'DE-120383D-A', 'DE-69840731-A']

```
#standardSQL
```

```
SELECT
```

```
  COUNT(*) AS cnt,
```

```
  COUNT(second.second_column) AS second_cnt,
```

```
  first.country_code AS grouped,
```

```
  ARRAY_AGG(first.application_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
```

```
FROM `patents-public-data.patents.publications` AS first
```

```

LEFT JOIN (
  SELECT application_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.uspto_oce_office_actions.match_app`
  GROUP BY 1
) AS second ON first.application_number = second.second_column
GROUP BY 3

```

joins to patents-public-data:uspto_oce_pair.match::application_number on application_number (11.95%, 11,727,859 rows)

Key	Percent	Rows	Sample values
CA	none	0	['CA-781516D-A', '', 'CA-2224077-A', 'CA-736648D-A', 'CA-179534D-A']
FR	none	0	['FR-1552705D-A', 'FR-894008-A', 'FR-0857142-A', 'FR-819852-A', 'FR-535689D-A']
RU	none	0	['SU-4911241-A', 'RU-2012116809-A', 'RU-2000116034-U', 'RU-99113495-A', 'RU-2013143163-A']
WO	29.63%	1,201,246	['AT-9200167-W', 'US-2006006348-W', 'US-2011045025-W', 'RU-2016000657-W', 'IL-0000073-W']
US	65.44%	10,526,613	['US-29232505-A', 'US-87239107-A', 'US-25132505-A', 'US-19746288-A', 'US-57213195-A']
BE	none	0	['BE-155177-A', 'BE-185049-A', '', '', 'BE-366825D-A']
DE	none	0	['DE-2516089-A', 'DE-3601193-A', 'DE-69601417-T', 'DE-59809022-A', '']
DK	none	0	['DK-199487-A', 'DK-302678-A', 'DK-01918032-T', 'DK-44020D-A', 'DK-1094D-A']
ES	none	0	['ES-93911047-T', 'ES-550566-A', 'ES-209717-U', 'ES-215953-A', 'ES-88402167-T']
GB	none	0	['GB-8912188-A', 'GB-8406090-A', 'GB-1838233-A', 'GB-2591077-A', 'GB-9807617-A']
KR	none	0	['KR-900009805-U', 'KR-19970007777-A', 'KR-20137018309-A', 'KR-19990027088-A', 'KR-19950025819-U']
EP	none	0	['EP-02708527-A', 'EP-03016939-A', 'EP-79200311-A', 'EP-84113938-A', 'EP-88100399-A']
JP	none	0	['JP-22233295-A', 'JP-2010500047-A', 'JP-15184183-A', 'JP-4496779-A', 'JP-22030485-A']
FI	none	0	['FI-832722-A', 'FI-971148-A', '', 'FI-20061043-A', '']
NL	none	0	['NL-272747D-A', 'NL-7212760-A', 'NL-255567D-A', 'NL-7014499-A', 'NL-7903729-A']
CN	none	0	['CN-201610963148-A', 'CN-201320298516-U', 'CN-96198651-A', 'CN-00130775-A', 'CN-201220200123-U']
LU	none	0	['LU-31403D-A', 'LU-60858D-A', 'LU-51422-A', 'LU-58699D-A', 'LU-58512D-A']

```

#standardSQL
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.application_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT application_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.uspto_oce_pair.match`
  GROUP BY 1
) AS second ON first.application_number = second.second_column
GROUP BY 3

```

joins to patents-public-data:uspto_peds.match::application_number on application_number (13.86%,

13,603,596 rows)

Key	Percent	Rows	Sample values
CA	11.25%	357,636	['CA-2037874-A', 'CA-2087875-A', 'CA-325915-A', 'CA-44277D-A', 'CA-2368094-A']
FR	0.06%	1,835	['FR-840958-A', 'FR-666930D-A', 'FR-801118D-A', 'FR-0310343-A', 'FR-995148-A']
RU	0.01%	151	['RU-2014112976-U', 'RU-2006119015-A', 'RU-2004132347-A', 'RU-2013129907-U', 'RU-2016145330-A']
WO	32.45%	1,315,294	['IN-2006000486-W', 'GB-9701650-W', 'IB-2006052755-W', 'US-0331817-W', 'JP-2005020591-W']
US	70.81%	11,391,162	['US-10774049-A', 'US-55535700-A', 'US-55290290-A', 'US-80887607-A', 'US-77304107-A']
BE	26.09%	206,627	['BE-603437-A', 'BE-779932-A', 'BE-9500348-A', 'BE-150267-A', 'BE-725507D-A']
DE	0.40%	30,477	['DE-2747637-A', 'DE-F0003037-D', 'DE-102017009963-A', 'DE-R0021363-U', 'DE-69027462-A']
EP	0.00%	3	['EP-10728575-A', 'EP-14748938-A', 'EP-05257544-A', 'EP-00992901-A', 'EP-02804541-A']
ES	3.81%	62,794	['ES-03725310-T', 'ES-02707077-T', 'ES-10566-U', 'ES-02017258-T', 'ES-96908130-T']
GB	0.09%	3,170	['GB-9413434-A', 'GB-1623369-A', 'GB-8721560-A', 'GB-8332012-A', 'GB-9413727-A']
KR	0.00%	1	['KR-20090016138-A', 'KR-20110122194-A', 'KR-19990039999-A', 'KR-19980020427-A', 'KR-20050126774-A']
DK	6.00%	34,808	['DK-10570D-A', 'DK-08852142-T', 'DK-BA200300327-U', '', 'DK-141472-A']
JP	0.00%	11	['JP-2006339114-A', 'JP-27256688-A', 'JP-15821279-A', 'JP-17226783-U', 'JP-2005284467-A']
FI	28.25%	174,286	['FI-784046-A', '', '', 'FI-840536-A', 'FI-954248-A']
NL	1.17%	7,562	['NL-7706754-A', 'NL-86904D-A', 'NL-7003578-A', 'NL-268865-A', 'NL-6500587-A']
CN	0.01%	1,509	['CN-201610255860-A', 'CN-201410085792-A', 'CN-201610281247-A', 'CN-201310251728-A', 'CN-201510443004-A']
LU	20.83%	16,270	['LU-40770D-A', 'LU-40086D-A', 'LU-66059D-A', 'LU-54934D-A', 'LU-81308-A']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.application_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT application_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.uspto_peds.match`
  GROUP BY 1
) AS second ON first.application_number = second.second_column
GROUP BY 3

```

joins to patents-public-data:uspto_ptab.match::application_number on application_number (0.01%, 7,832 rows)

Key	Percent	Rows	Sample values
CA	<i>none</i>	0	['CA-17553D-A', 'CA-59698D-A', 'CA-2781514-A', 'CA-2198796-A', 'CA-2337801-A']

Key	Percent	Rows	Sample values
FR	<i>none</i>	0	['FR-9016242-A', 'FR-7201675-A', 'FR-1662221-A', 'FR-1017927D-A', 'FR-1028344D-A']
RU	<i>none</i>	0	['RU-2014123281-U', 'RU-2013147672-U', 'SU-4901723-A', 'RU-2009146588-A', 'RU-2003131054-A']
WO	<i>none</i>	0	['US-2007080800-W', 'US-2010049685-W', 'GB-0103139-W', 'US-2015025360-W', 'US-9500116-W']
US	0.05%	7,832	['US-201615041550-A', 'US-201314136269-A', 'US-201113193367-A', 'US-87123607-A', 'US-201113878426-A']
BE	<i>none</i>	0	['', 'BE-669566-A', 'BE-409721D-A', 'BE-558920D-A', '']
GB	<i>none</i>	0	['GB-189514570D-A', 'GB-201322995-A', 'GB-9022015-A', 'GB-8723854-A', 'GB-4096569-A']
NL	<i>none</i>	0	['NL-52595D-A', 'NL-7908304-A', 'NL-6618064-A', 'NL-267354D-A', 'NL-7104038-A']
DE	<i>none</i>	0	['DE-246869D-A', 'DE-2624061-A', 'DE-F0030030-U', 'DE-1604434-A', 'DE-59601455-A']
LU	<i>none</i>	0	['LU-29705D-A', 'LU-66572D-A', 'LU-52161D-A', 'LU-75653-A', 'LU-39975D-A']
DK	<i>none</i>	0	['DK-91900101-T', 'DK-343883-A', 'DK-93905439-T', 'DK-99915479-T', '']
KR	<i>none</i>	0	['KR-20030083845-A', 'KR-20090083002-A', 'KR-20117003818-A', 'KR-20167033768-A', 'KR-20140147727-A']
EP	<i>none</i>	0	['EP-88112017-A', 'EP-13796228-A', 'EP-08774951-A', 'EP-02758823-A', 'EP-90124106-A']
JP	<i>none</i>	0	['JP-12707675-A', 'JP-5689275-U', 'JP-41011390-A', 'JP-17598777-U', 'JP-10849672-U']
FI	<i>none</i>	0	['', 'FI-991199-A', 'FI-U980295-U', 'FI-110870-A']
ES	<i>none</i>	0	['ES-05292307-T', 'ES-257859D-A', 'ES-534786-A', 'ES-548504-A', 'ES-03702481-T']
CN	<i>none</i>	0	['CN-201610527052-A', 'CN-201710239071-A', 'CN-201711368702-A', 'CN-201320005431-U', 'CN-201710165702-A']

#standardSQL

SELECT

COUNT(*) AS cnt,

COUNT(second.second_column) AS second_cnt,

first.country_code AS grouped,

ARRAY_AGG(first.application_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value

FROM `patents-public-data.patents.publications` AS first

LEFT JOIN (

SELECT application_number AS second_column, COUNT(*) AS cnt

FROM `patents-public-data.uspto_ptab.match`

GROUP BY 1

) AS second ON first.application_number = second.second_column

GROUP BY 3

joins from patents-public-data:uspto_oce_office_actions.match_app::application_number on application_number (99.80%, 2,183,770 rows)

joins from patents-public-data:uspto_oce_pair.match::application_number on application_number (100.00%, 8,437,907 rows)

joins from patents-public-data:uspto_peds.match::application_number on application_number (100.00%, 9,162,229 rows)

joins from patents-public-data:uspto_ptab.match::application_number on application_number (100.00%, 4,610 rows)

family_id

joins to patents-public-data:uspto_oce_cancer.publications::Family_ID on family_id (1.07%, 1,047,661 rows)

Key	Percent	Rows	Sample values
CA	2.51%	79,878	['22339137', '47553798', '43875627', '27328926', '35972487']
FR	0.24%	7,552	['9592731', '47902098', '1376007', '9429902', '9555177']
RU	1.36%	17,095	['21546221', '20155868', '44803738', '21541030', '44996692']
WO	3.58%	145,179	['27511480', '39345280', '52273075', '40526883', '35784379']
US	1.81%	291,814	['48523668', '23103645', '52630340', '38917819', '25340171']
BE	0.16%	1,259	['26728796', '25112275', '21988951', '186781', '32966128']
LU	0.74%	575	['25656021', '19730470', '1824483', '19728767', '11160661']
DK	3.46%	20,087	['7911677', '20343630', '6277804', '61274984', '27021562']
ES	1.70%	28,024	['8340536', '38610000', '34495574', '10845985', '8258938']
GB	0.29%	10,855	['10585558', '10593560', '49767396', '26264062', '13722883']
KR	0.75%	36,243	['51999097', '32923783', '55582796', '42754657', '40697366']
EP	2.92%	183,528	['49116910', '6474027', '6313727', '35786493', '38859081']
JP	0.47%	115,038	['31317615', '26548330', '11898182', '28834875', '32024094']
FI	1.41%	8,676	['26159387', '47915285', '27010156', '26657748', '10553181']
NL	0.27%	1,735	['1878845', '5937684', '5871772', '6932278', '5852142']
CN	0.31%	60,213	['40100867', '51499059', '55379166', '39891457', '18685794']
DE	0.52%	39,910	['26253592', '34353292', '25830299', '465631', '730556']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.family_id IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT Family_ID AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.uspto_oce_cancer.publications`
  GROUP BY 1
) AS second ON first.family_id = second.second_column
GROUP BY 3

```

joins to patents-public-data:dsep.disclosures_13::family_id on family_id (0.06%, 61,139 rows)

Key	Percent	Rows	Sample values
CA	0.11%	3,572	['36411480', '36303238', '8226366', '35606721', '25322276']
FR	0.01%	400	['9278953', '8855634', '8760618', '16003986', '9292304']
RU	0.10%	1,220	['21584557', '21564142', '37059463', '20159436', '49182855']
WO	0.11%	4,454	['12597157', '38291300', '35395222', '59623495', '20408528']
US	0.08%	13,258	['22769984', '38520602', '12943837', '6910943', '32376214']
BE	0.00%	33	['25648590', '24689822', '24592587', '9092924', '30773896']
GB	0.02%	792	['24962476', '10363824', '10179478', '6635885', '7095733']
NL	0.01%	89	['1868026', '1821574', '19852173', '24235196', '7263240']
ES	0.10%	1,718	['37636756', '37843032', '34629609', '7808045', '8367173']
LU	0.00%	1	['26325000', '1768159', '19731128', '3865254', '5767944']
DK	0.15%	842	['26975113', '19774634', '10479160', '6364379', '25436362']
KR	0.08%	3,831	['49381658', '54833763', '46715786', '41502797', '39265188']
EP	0.18%	11,115	['44120974', '18449055', '27758892', '7921086', '51790599']
JP	0.03%	6,918	['11776381', '35731577', '30372053', '31744099', '28733122']
FI	0.39%	2,418	['7096533', '9228307', '8545325', '20397011', '8531063']
CN	0.03%	5,754	['57566123', '47645410', '55312380', '52038935', '56347128']
DE	0.06%	4,724	['17966130', '7047681', '22314741', '6075770', '48742094']

#standardSQL

```

SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.family_id IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT family_id AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.dsep.disclosures_13`
  GROUP BY 1
) AS second ON first.family_id = second.second_column
GROUP BY 3

```

joins to erudite-marker-539:JEMS16.patent_metadata_2::FamilyID on **family_id** (25.67%, 25,206,642 rows)

Key	Percent	Rows	Sample values
CA	39.58%	1,257,863	['4227856', '24699995', '9930135', '48953906', '38227680']
FR	15.60%	491,090	['8578730', '8993842', '6677627', '11139482', '8988220']
RU	14.33%	180,633	['45406150', '20248686', '50666248', '21454265', '60328542']
WO	37.76%	1,530,699	['39626164', '51260886', '45481822', '37762525', '38923074']
US	55.32%	8,899,874	['17141468', '24795238', '57451223', '24855630', '40939755']
BE	11.25%	89,086	['27186844', '148199', '27365', '98944', '23087143']
DE	24.75%	1,886,290	['25909863', '8020775', '7380137', '7970920', '53484878']
DK	40.74%	236,260	['9127503', '5901894', '8146592', '5634272', '5901321']
ES	27.36%	451,162	['8211140', '19798077', '34638566', '8377218', '8291221']
EP	56.41%	3,551,033	['38801720', '26344324', '40259347', '34135978', '23771646']
LU	13.08%	10,216	['9694031', '7106247', '5942648', '21860425', '1766308']
KR	19.30%	928,458	['38158704', '26628508', '45028202', '38924371', '19559145']
NL	16.05%	103,449	['1832780', '1802947', '27544899', '19857662', '25317719']
JP	12.28%	2,998,218	['29297687', '43040491', '24567510', '13976585', '18259620']
GB	18.89%	703,021	['10214017', '25054828', '35248534', '22738001', '10505611']
FI	36.42%	224,751	['19727442', '8537782', '6103458', '36876801', '8542204']
CN	8.66%	1,664,539	['61871096', '53817674', '41469781', '59487432', '53227073']

```

#standardSQL
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  first.country_code AS grouped,
  ARRAY_AGG(first.family_id IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.patents.publications` AS first
LEFT JOIN (
  SELECT FamilyID AS second_column, COUNT(*) AS cnt
  FROM `erudite-marker-539.JEMS16.patent_metadata_2`
  GROUP BY 1
) AS second ON first.family_id = second.second_column
GROUP BY 3

```

joins from patents-public-data:uspto_oce_cancer.publications::Family_ID on **family_id** (99.05%, 266,792 rows)

joins from patents-public-data:dsep.disclosures_13::family_id on **family_id** (29.46%, 13,358 rows)

joins from erudite-marker-539:JEMS16.patent_metadata_2::FamilyID on **family_id** (87.45%, 5,677,428 rows)

patents-public-data:patents.publications__201710

Old table version 201710, schema skipped.

patents-public-data:patents.publications_201802

Old table version 201802, schema skipped.

patents-public-data:google_patents_research.publications

Google Patents Research Data contains the output of much of the data analysis work used in Google Patents (patents.google.com), including machine translations of titles and abstracts from Google Translate, extracted top terms, similar documents, and forward references.

“Google Patents Research Data” by Google, based on data provided by IFI CLAIMS Patent Services, is licensed under a Creative Commons Attribution 4.0 International License.

Stat	Value
Last updated	2018-11-26
Rows	98,176,830
Size	254.2 GB

Schema

View in BigQuery

- **publication_number** STRING NULLABLE joins on **publication_number**
Patent publication number (DOCDB compatible), eg: ‘US-7650331-B1’.
- **title** STRING NULLABLE
The (possibly machine translated) English title.
- **title_translated** BOOLEAN NULLABLE
True if the title is machine-translated by Google Translate.
- **abstract** STRING NULLABLE
The (possibly machine translated) English abstract.
- **abstract_translated** BOOLEAN NULLABLE
True if the abstract is machine-translated by Google Translate.
- **cpc** RECORD REPEATED
The Cooperative Patent Classification (CPC) codes with the hierarchy.
- **cpc.code** STRING NULLABLE
Classification code
- **cpc.inventive** BOOLEAN NULLABLE
Is this classification inventive/main?
- **cpc.first** BOOLEAN NULLABLE
Is this classification the first/primary?
- **cpc.tree** STRING REPEATED
The full classification tree from the root to this code
- **cpc_low** STRING REPEATED

The Cooperative Patent Classification (CPC) codes and their parents, in an array for easier querying.

- `cpc_inventive_low` STRING REPEATED

The inventive Cooperative Patent Classification (CPC) codes and their parents, in an array for easier querying.

- `top_terms` STRING REPEATED

The top 10 salient terms extracted from this patent's title, abstract, claims and description.

- `similar` RECORD REPEATED

Semantically similar documents based on content and metadata.

- `similar.publication_number` STRING NULLABLE

Same as `publication_number`

- `similar.application_number` STRING NULLABLE

Same as `application_number`

- `similar.npl_text` STRING NULLABLE

Free-text citation (non-patent literature, etc).

- `similar.type` STRING NULLABLE

The type of reference (see parent field for values).

- `similar.category` STRING NULLABLE

The category of reference (see parent field for values).

- `similar.filing_date` INTEGER NULLABLE

The filing date.

- `url` STRING NULLABLE

URL to the `patents.google.com` page for this patent.

- `country` STRING NULLABLE

Country name.

- `publication_description` STRING NULLABLE

Description of the publication type.

- `cited_by` RECORD REPEATED

Publications that cite this publication.

- `cited_by.publication_number` STRING NULLABLE

Same as `publication_number`

- `cited_by.application_number` STRING NULLABLE

Same as `application_number`

- `cited_by.npl_text` STRING NULLABLE

Free-text citation (non-patent literature, etc).

- `cited_by.type` STRING NULLABLE

The type of reference (see parent field for values).

- `cited_by.category` STRING NULLABLE

The category of reference (see parent field for values).

- `cited_by.filing_date` INTEGER NULLABLE

The filing date.

- **embedding_v1** FLOAT REPEATED

(Version 1) Machine-learned vector embedding based on document contents and metadata, where two documents that have similar technical content have a high dot product score of their embedding vectors. Trained on full text bag of words to predict CPCs using the WSABIE classification model.

Join columns

publication__number

joins to patents-public-data:patents.publications::publication_number on **publication__number** (100.00%, 98,176,830 rows)

Key	Percent	Rows	Sample values
all	100.00%	98,176,830	['NL-6602752-A', 'CN-103831460-A', 'JP-6162077-B2', 'US-5859580-A', 'EP-1990470-A2']

#standardSQL

```
SELECT
  COUNT(*) AS cnt,
  COUNT(second.second_column) AS second_cnt,
  ARRAY_AGG(first.publication_number IGNORE NULLS ORDER BY RAND() LIMIT 5) AS sample_value
FROM `patents-public-data.google_patents_research.publications` AS first
LEFT JOIN (
  SELECT publication_number AS second_column, COUNT(*) AS cnt
  FROM `patents-public-data.patents.publications`
  GROUP BY 1
) AS second ON first.publication_number = second.second_column
```

joins from patents-public-data:patents.publications::publication_number on **publication__number** (100.00%, 98,176,830 rows)

patents-public-data:google_patents_research.publications_201710

Old table version 201710, schema skipped.

patents-public-data:google_patents_research.publications_201802

Old table version 201802, schema skipped.