Patents

Trade Marks

Designs

Plant Breeder's Rights

Delivering a world leading IP system

Intellectual Property Longitudinal Research Data (IPLoRD2020) Data Dictionary

Office of the Chief Data Officer

April 2021

Contents

What is IPLoRD?

IPLoRD – Intellectual Property Longitudinal Research Dataset – is a publicly available dataset that provides researchers with an applicant activity-centric view of Australian IP right filings. IPLoRD leverages IPGOD data to provide a history of an applicant's filings over time broken down by financial year. Where available, the ABN of the applicant is provided to allow linkage with other government research datasets.

IPLoRD makes it possible for economists and other researchers to examine relationships between the filer's economic circumstances and their intellectual property portfolio.

IPLoRD is a biennial release. This year's version of IPLoRD provides coverage of filing activities from 1904 to the end of 2020.

What's new in IPLoRD2020

IPLoRD has undergone some changes since its inaugural release in 2018.

Along with IPGOD2020, IPPLoRD has undergone some changes since its inaugural release in 2018, and it has been re-engineered from source to presentation. This enhancement enables provision of additional data fields that were not available in the previous version, particularly around classification of filing activity. We are now providing:

- Information about the type of product or service associated with a trade mark as well as classification information about design and plant breeder's right filings. Coverage of patent classification information has been maintained.

- Additional information regarding the most common (modal) classification areas associated with an applicant's activity.

IPLoRD retains the same general structure as previous releases, including information regarding filing, granting, retiring and stockpiles of IP right filings as well as information about applicants, including (where available) the ABN of the business filing for an IP Right.

Getting in touch

We are committed to making IPLoRD as usable, comprehensive, and accurate as possible. So, we want to hear from you to ensure the product continues to meet your needs. If you have feedback or comments, please contact us via the data.gov.au portal.

IPLoRD Structure

IPLoRD is a longitudinal research table that provides information in a wide format to enable easy reference to the IP activity of applicants in Australia across time. The table is intended to facilitate use by users from many technical backgrounds and has been structured to ensure user-friendliness.

Variable Categories

IPLoRD includes the following categories of variables:

1. **Identifying variables**

2. **Activity variables**

Identifying variables are chiefly concerned with high-level information about the right applicant. IPLoRD uses IPGOD2020 as a source and utilises ABN values obtained from the Australian Business Register. This information and key location information is provided to help identify an applicant and localise their activity.

The Australian financial year is the financial year for which the activity variables are calculated.

The activity metric variables describe the activity of applicants in different ways to indicate the quantity of activity and the type of activities being conducted.

Identifying Variables

The identifying variables in IPLoRD are described below:

Variable	Definition
lord_party_id	Unique identifier for the party in IPLoRD. Depending on the value of lord_party_id_type it is either an ABN or an IPGOD2020 party identifier
lord_party_id_type	One of: - abn – indicating that the id variable is the ABN - party_id – indicates the lord_party_id is the party_id as it appears in the party-activity table of IPGOD2020. ABN is used preferentially as it provides the best linkage to other Australian economic research data sets. In many cases, we have provided the ABN of the Australian subsidiary for international firms in the dataset.
country_code	A two-character code representing the identified country of the filing activity. IPGOD and its sources are not static data products and individuals and firms may move location or conduct activities from distinct locations while remaining the same entity across time. IPLoRD uses the modal country code for each lord_party_id where there is a conflict.
state_code	Australian state code of the applicant As for country_code, modal values are taken if the applicant has multiple associated state codes.
post_code	The post code of the applicant. These may be both Australian post codes and post codes from other jurisdictions. As for country_code, modal values are used where more than one post_code is identified as being associated with an applicant.

In this new release, we also included a range of variables that identifies when an applicant first applied for or was granted a particular type of IP right.

These fields are labelled 'first' and include:

p_first_application_financial_year
p_first_gained_enforceable_status_financial_year
pbr_first_application_financial_year
pbr_first_gained_enforceable_status_financial_year
tm_first_application_financial_year
tm_first_gained_enforceable_status_financial_year
d_first_application_financial_year
d_first_gained_enforceable_status_financial_year

Gained Enforceable Status is used in the IPGOD dataset to describe when a right is granted legally enforceable status by IP Australia. The terminology is reproduced here for simpler interoperability with the parent dataset.

Activity variables

Activity variables use activity metrics to describe what happened to an applicant's IP right portfolio in a given financial year across each of the event streams. Activity metrics are measures of filing activity from the perspective of volume, diversity, or area of innovation. Each of the activity metrics and event streams is defined and described in further detail below.

Count variables

Count variables are counts of the number of events of a certain kind or the number of rights in the stockpile of an applicant for an IP right.

Weighted count variables

Weighted count variables consider the number of applicants associated with an IP Right. For example, when an application is filed jointly by a group of **n** applicants, that application counts (**1/n**) towards each of the applicants' weighted count score. The inclusion of weighted count variables allows for accurate numbers of applications to be preserved in aggregate for each IP right type.

Modal value variables

Modal value variables record the most frequent variable type (using IPLoRD's simplified classification system for IP activity) for activity in a year. They are useful for getting a high-level understanding of an applicant's IP filing's economic area and the main focuses of the company's activity.

Scope classification variables

Scope variables are intended to indicate the breadth of an applicant's activity across multiple event streams in a given financial year and are used as a term to describe the number of classification areas in which an applicant conducts the activity in a given financial year.

For example, an applicant with a trademark granting scope of 20 in 2019/20 would have had trademark applications granted with classifications in 20 different trademark areas (of the possible 45 in the Nice

system). This indicates economic activity associated with a broader range of products than an applicant with a smaller scope.

Each of the activity metrics - count, weighted count, modal class and scope class variables can be considered about each of the event streams described below and, where possible, within different coarse classification areas of filing activity for patents and trade marks.

IP Right classification simplification

For the activity metrics classification scope and modal classification, it is necessary to significantly simplify or modify the classification system used by the original right system.

To describe filing activity by subject matter, IP right systems use a classification standard:

- Patents has the International Patent Classification (IPC) system.

- Trademarks uses the Nice classification system to categorise marks as being associated with certain types of goods or services.

- The Locarno system describes product types associated with protected Design rights (industrial design patents in other jurisdictions).

- A Plant Breeder's right can be classified by the plant's genus and species being protected.

These classification systems can provide exact indications of the content of an IP right filing. Particularly in patents and trade marks, it is common to have multiple classifications assigned to them, allowing the IP Right to be described in detail. While in designs, multiple classification types for a single Right are possible, they are rare.

Classification area & coarse classification area

IPLoRD employs two distinct concepts of classification:

- classification area. This is used for all 4 right types, and is either a simplification or a reproduction of the right-specific classification for the right types; and

- coarse classification area, which is an aggregation of classification areas for patents and trade marks.

These levels of aggregation facilitate the creation of a range of variables for specific kinds of analysis. These classification schemes are described below.

Classification area

For patents, designs and trade marks there are 35, 32 and 45 possible classification areas respectively.

Classification area is used as the variable of interest (i.e. the variable counted over or selected) for modal class and class scope variables. As IPLoRD is intended as a table for high-level economic analysis of an organisation's IP right filing activity across time, a simpler classification system is required for some rights.

The Locarno system for designs has 32 high-level categories (classification areas). The first two digits of the Locarno symbol determine which high-level group the right belongs to. These are described in WIPO literature.

Similarly, the Nice system for trade marks has 45 possible classification areas. These have been left unaltered for IPLoRD. The meaning of these can be found from WIPO literature.

Simplifying PBR Classifications

The source dataset for IPLoRD is IPGOD2020 which uses genus and species in combination to describe the right filing. In IPLoRD this has been simplified to simply the genus.

Simplifying IPC (Patent) Classification

The IPC classification system comprises tens of thousands of classification symbols which may be applied either singly or multiply to a patent application to describe its subject matter. The IPC is designed for precise searching by patent examiners to determine novelty and inventive step. They are not intended for use by lay audiences to gain a high-level understanding of a company's IP activity.

This problem is not novel and has been, to a large extent, solved by the World Intellectual Property Organisation in their [technology to patent classification concordance for country comparison](). Table 2 of that document describes how to use IPC marks to assign a patent to one or more of 35 possible technology areas. This 35 category classification system has been used for IPLoRD. It is possible to have more than one of these classifications associated with a right.

The IPC is not a static standard, with new technologies adding to and changing the classification system. Since 2008, when the WIPO IPC tech area concordance was written, several new classification symbols have emerged. These have been manually added to the original standard. The simplified patent classification system as modified for IPLoRD production is included in Table 1 in Annex A – Simplified Patent Classification System.

Coarse classification area

For patents and trade marks specifically, it's possible to talk about a higher-level of classification, which we have termed 'coarse classification area'. A coarse classification area is a set of groups made up of different classification areas which can be meaningfully used to break activity for a given right type into a handful of groups for analysis at a more granular level than looking at the right type as a whole. Modal class, class scope and count variables are provided in IPLoRD both for the rights as a whole and for rights within each coarse classification group.

Patents the coarse classification areas comprise five groupings of classification areas:

- Electrical Engineering (8 classification areas)
- Chemistry (11 classification areas)
- Instruments (5 classification areas)
- Mechanical Engineering (8 classification areas)
- Other (3 classification areas)

For further information, consult Table 1 in Annex A – Simplified Patent Classification System. For patents, it is possible for a right to exist in more than one of the coarse classification areas simultaneously.

Trade marks there are three possible coarse classification areas:

- Goods: having only Nice classifications 1-34
- Services: having only Nice classifications 35-45
- Goods and Services: having at least one goods and one services classification symbol.

For trade marks ,coarse classification areas should be non-overlapping (disjoint) sets.

Event Streams

IPLoRD tracks the activity metrics across several different event types (or event streams) in an IP right's lifecycle. The event streams considered are:

- Filed – rights filed during the financial year
- Granted – rights granted during the financial year
- Retired – rights retired during the financial year

- Enforceable – Stockpile of granted, non-retired rights at the end of each financial year

The enforceable stockpile is affected by flows from the granted and retired applications. Each year's enforceable stockpile will be the previous year's enforceable stockpile plus rights granted in that financial year and rights retired in that financial year. Filings do not affect the stockpile.

Naming Conventions for non-identity variables

The naming conventions for non-identity variables are as follows:

<ip_right_type>_<event_stream>_<activity_metric >

IP right types are abbreviated as:

- p – patents
- d – designs
- tm – trade marks
- pbr – plant breeder's rights

Therefore, the count (activity_metric) of the enforceable stockpile (event stream) of PBRs is recorded as pbr_enforceable_count.

Similarly, the modal classification (activity_metric) of trade marks retired (event stream) is recorded as the variable tm_retired_modal_class.

In the case of trade marks and patents, the rights may be subsetted by coarse classification area into several smaller groups according to the right's subject matter. This is represented in the data in the following way:

<ip_right_type>_<coarse _classification_area>_<event_stream>_<activity_metric >

For example, if we are interested to know how many of the eight possible electrical engineering classification areas a firm has had patents granted in a given financial year, we would look for p_electrical_engineering_granted_class_scope. Similarly, a firm's count of goods trade marks filed each year would be tm_goods_filed_count.

In this way, it is possible to understand the meaning of each of the fields produced in IPLoRD without the need to describe all the variables individually.

Using this system of variable name description, IPLoRD can provide 151 variables that describe the areas of activity for each of the approximate 900,000 distinct applicants IP Australia has interacted with across its 116-year history.

Annex A – Simplified Patent Classification System

Table 1 - Technology Area to IPC Concordance - Reproduced from WIPO with Bold Terms Added to Align with IPC Classification Changes

Technology Grouping (Coarse Classification Area)	Technology Area (Classification Area)	IPC Symbol Prefixes
Electrical Engineering	1 Electrical machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
Electrical Engineering	2 Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K, **H04N-019**
Electrical Engineering	3 Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q, **H04N-101**
Electrical Engineering	4 Digital communication	H04L, **H04W, H04N-021**
Electrical Engineering	5 Basic communication processes	H03#
Electrical Engineering	6 Computer technology	(G06# not G06Q), G11C, G10L, **G16C, G16B, G16Y, G16Z**
Electrical Engineering	7 IT methods for management	G06Q
Electrical Engineering	8 Semiconductors	H01L
Instruments	9 Optics	G02#, G03B, G03C, G03D, G03F, G03G, G03H, H01S
Instruments	10 Measurement	(G01N not G01N-033), G01P, G01R, G01S; G01V, G01W, G04#, G12B, G99Z, **G01Q**
Instruments	11 Analysis of biological materials	G01N-033
Instruments	12 Control	G05B, G05D, G05F, G07#, G08B, G08G, G09B, G09C, G09D
Instruments	13 Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G, **G16H**

Technology Grouping (Coarse Classification Area)	Technology Area (Classification Area)	IPC Symbol Prefixes
Chemistry	14 Organic fine chemistry	(C07B, C07C, C07D, C07F, C07H, C07J, C40B) not A61K, A61K-008, A61Q
Chemistry	15 Biotechnology	(C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S) not A61K
Chemistry	16 Pharmaceuticals	A61K not A61K-008, A61P
Chemistry	17 Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
Chemistry	18 Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K, **C13B-010, C13B-020, C13B-030, C13B-35, C13B-040, C13B-050, C13B-099**
Chemistry	19 Basic materials chemistry	A01N, A01P, C05#, C06#, C09B, C09C, C09F, C09G, C09H, C09K, C09D, C09J, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
Chemistry	20 Materials, metallurgy	C01#, C03C, C04#, C21#, C22#, B22#
Chemistry	21 Surface technology, coating	B05C, B05D, B32#, C23#, C25#, C30#
Chemistry	22 Micro-structure and nano-technology	B81#, B82#
Chemistry	23 Chemical engineering	B01B, B01D-000#, B01D-01##, B01D-02##, B01D-03##, B01D-041, B01D-043, B01D-057, B01D-059, B01D-06##, B01D-07##, B01F, B01J, B01L, B02C, B03#, B04#, B05B, B06B, B07#, B08#, D06B, D06C, D06L, F25J, F26#, C14C, H05H
Chemistry	24 Environmental technology	A62D, B01D-045, B01D-046, B01D-047, B01D-049, B01D-050, B01D051, B01D-052, B01D-053,

Technology Grouping (Coarse Classification Area)	Technology Area (Classification Area)	IPC Symbol Prefixes
		B09#, B65F, C02#, F01N, F23G, F23J, G01T, E01F-008, A62C
Mechanical Engineering	25 Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66#, B67#
Mechanical Engineering	26 Machine tools	B21#, B23#, B24#, B26D, B26F, B27#, B30#, B25B, B25C, B25D, B25F, B25G, B25H, B26B
Mechanical Engineering	27 Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02#, F03#, F04#, F23R, G21#, F99Z
Mechanical Engineering	28 Textile and paper machines	A41H, A43D, A46D, C14B, D01#, D02#, D03#, D04B, D04C, D04G, D04H, D05#, D06G, D06H, D06J, D06M, D06P, D06Q, D99Z, B31#, D21#, B41#
Mechanical Engineering	29 Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22#, A23N, A23P, B02B, C12L, C13C, C13G, C13H, B28#, B29#, C03B, C08J, B99Z, F41#, F42#, **C13B-005, C13B-015, C13B-025, C13B-045, B33Y**
Mechanical Engineering	30 Thermal processes and apparatus	F22#, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24#, F25B, F25C, F27#, F28#
Mechanical Engineering	31 Mechanical elements	F15#, F16#, F17#, G05G
Mechanical Engineering	32 Transport	B60#, B61#, B62#, B63B, B63C, B63G, B63H, B63J, B64#
Other Fields	33 Furniture, games	A47#, A63#
Other Fields	34 Other consumer goods	A24#, A41B, A41C, A41D, A41F, A41G, A42#, A43B, A43C, A44#, A45#, A46B, A62B, B42#, B43#, D04D, D07#, G10B, G10C, G10D, G10F, G10G, G10H, G10K, B44#, B68#, D06F, D06N, F25D, A99Z
Other Fields	35 Civil engineering	E02#, E01B, E01C, E01D, E01F-001, E01F-003, E01F-005, E01F-007, E01F-009, E01F-01#, E01H,

Technology Grouping (Coarse Classification Area)	Technology Area (Classification Area)	IPC Symbol Prefixes
		E03#, E04#, E05#, E06#, E21#, E99Z