Topic 6:

Free and Open Source Tools for Patent Analytics with reference to the WIPO Manual

Irene Kitsara, IP Information Officer,

Access to Information and Knowledge Division

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Tools for PLR Preparation

■ Efforts to create online semi-automated tools for the preparation of (semi-) automated PLR, e.g.

Tools for PLR preparation (2)
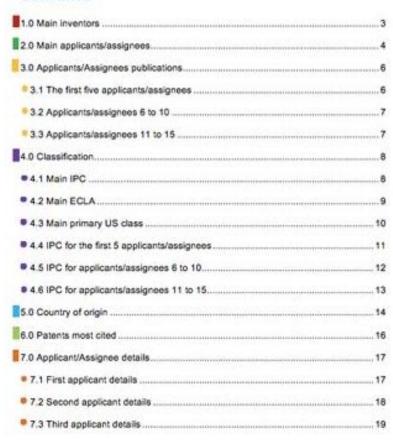
Social Media – Linkedin Groups on landscaping

- Patent Searching and Landscaping - http://www.linkedin.com/groups?home=&gid=110874&trk=anet_ug_hm

- IP Intelligence - http://www.linkedin.com/groups?gid=124885&trk=myg_ugrp_ovr

- Patinformatics - http://www.linkedin.com/groups?gid=2391676&trk=myg_ugrp_ovr

- Patent Landscaping Innovations - http://www.linkedin.com/groups/Patent-Landscaping-Innovations-1883637?trk=myg_ugrp_ovr

- Patent Analysis & Landscaping - http://www.linkedin.com/groups/Patent-Analysis-Landscaping-1794428?trk=myg_ugrp_ovr

- Patent Informatics Group - http://www.linkedin.com/groups?gid=76745&trk=myg_ugrp_ovr

- Patent Landscapes - http://www.linkedin.com/groups?gid=2846540&trk=myg_ugrp_ovr

- Competitive Intelligence from Patent Analytics - http://www.linkedin.com/groups/Competitive-intelligence-from-patent-analytics-3891330?trk=myg_ugrp_ovr
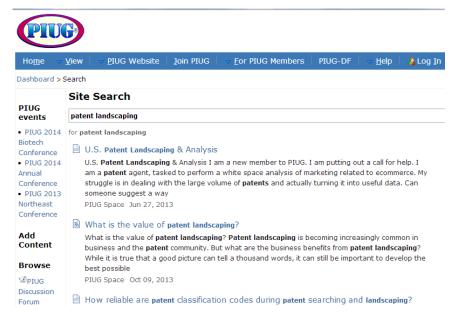
Blogs/Wiki/Online Fora on Patent Landscaping

www.patinformatics.com

http://wiki.piug.org

What type of tools are used in patent analytics?

- To help in various stages and tasks involved in patent analytics:
    - (Searching)
    - Obtaining
    - Cleaning/Normalizing/Tidying
    - Grouping/Clustering   Data
    - Analyzing
    - Visualizing
    - Sharing

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Tools used in Patent Analytics

■ Commercial:

　■ Embedded in patent databases

　■ Complementary to patent search tools and compatible

　■ Typically dedicated to patent data, with some exceptions

STN ANAVIST™　Questel IP Business Intelligence　THOMSON INNOVATION　PatBase ANALYTICS　TREPAREL POWERED BY EVALUESERVE

Relecura　INTELLIXIR　vantagepoint Search Technology

■ Free of charge:

　■ Some possibilities

PATENTSCOPE

OPEN Refine

OECD: HAN database
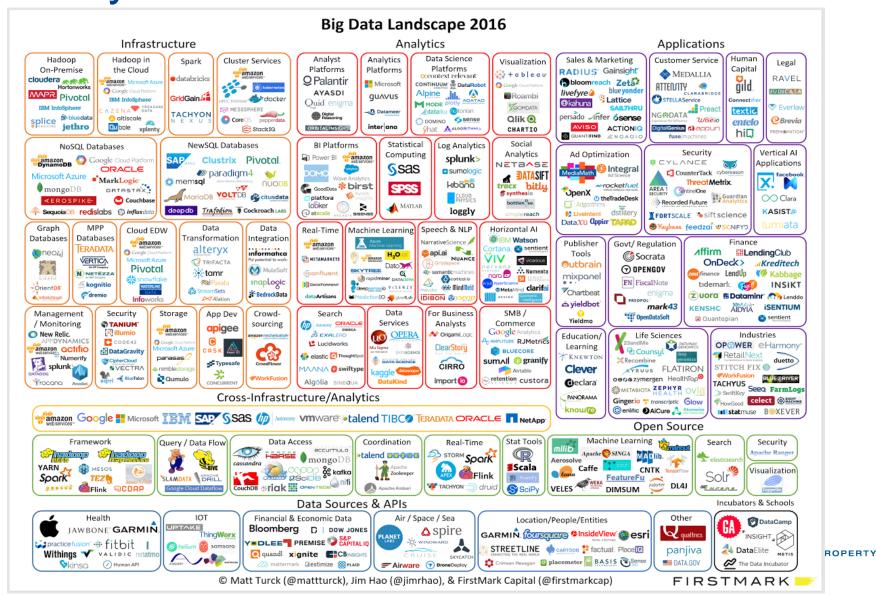
■ Open Source Tools: recent trend

(…some of the) tools facilitating data analysis

Big Data Landscape 2016

© Matt Turck (@mattturck), Jim Hao (@jimrhao), & FirstMark Capital (@firstmarkcap)
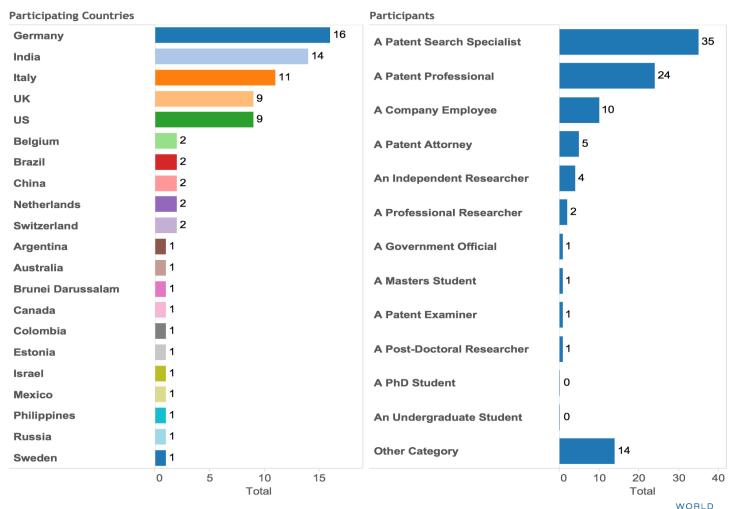
Which tools are used in patent analytics?

WIPO Survey on the use of open source and free tools for patent analytics

- Survey addressed to the community of patent information users on the tools they are using for patent analysis

- 13 questions in SurveyMonkey

- Shared through PIUG Wiki, the German and the Italian Patent Information User Groups, WIPO Twitter and PLR website, Linkedin Patent Landscape Group.

- Received 131 answers (81 answered all questions)

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Survey results: participants profile
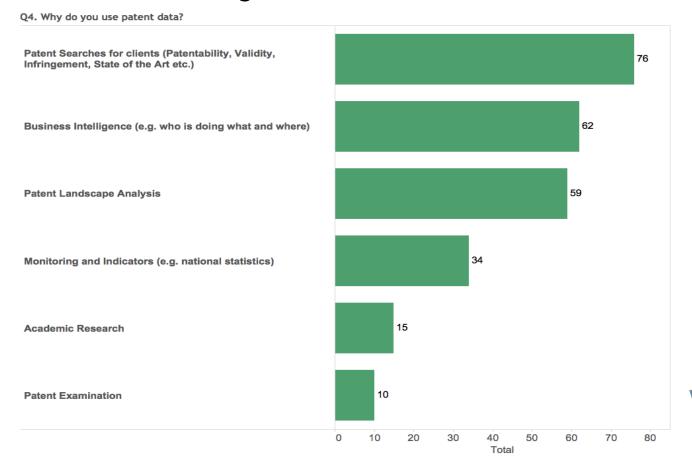
■ 131 answers from 21 countries, mainly patent search professionals

Participating Countries

Country	Total
Germany	16
India	14
Italy	11
UK	9
US	9
Belgium	2
Brazil	2
China	2
Netherlands	2
Switzerland	2
Argentina	1
Australia	1
Brunei Darussalam	1
Canada	1
Colombia	1
Estonia	1
Israel	1
Mexico	1
Philippines	1
Russia	1
Sweden	1

Participants

Category	Total
A Patent Search Specialist	35
A Patent Professional	24
A Company Employee	10
A Patent Attorney	5
An Independent Researcher	4
A Professional Researcher	2
A Government Official	1
A Masters Student	1
A Patent Examiner	1
A Post-Doctoral Researcher	1
A PhD Student	0
An Undergraduate Student	0
Other Category	14

WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Survey results: what do participants use analytics for?

- Mainly for clients – a lot carrying out landscaping or business intelligence

Q4. Why do you use patent data?

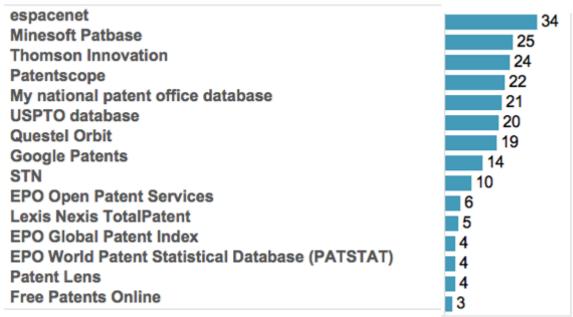| Category	Total
Patent Searches for clients (Patentability, Validity, Infringement, State of the Art etc.)	76
Business Intelligence (e.g. who is doing what and where)	62
Patent Landscape Analysis	59
Monitoring and Indicators (e.g. national statistics)	34
Academic Research	15
Patent Examination	10

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

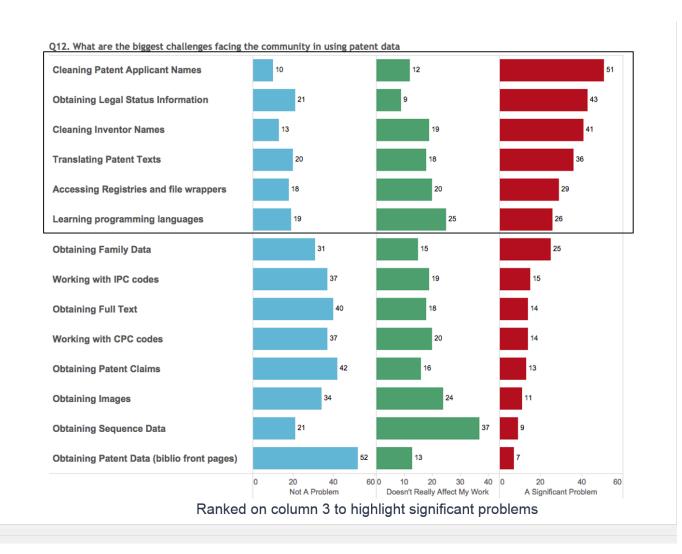# Survey results: Databases used for obtaining data for analytics

■ Mixture of free and commercial databases

■ Free databases: espacenet, followed by PATENTSCOPE

■ Commercial: Patbase, Thomson Innovation, Questel Orbit

Q5. What Patent Databases do you use?

Database	Count
espacenet	34
Minesoft Patbase	25
Thomson Innovation	24
Patentscope	22
My national patent office database	21
USPTO database	20
Questel Orbit	19
Google Patents	14
STN	10
EPO Open Patent Services	6
Lexis Nexis TotalPatent	5
EPO Global Patent Index	4
EPO World Patent Statistical Database (PATSTAT)	4
Patent Lens	4
Free Patents Online	3

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

Greatest challenges in patent analytics identified by the survey participants

Q12. What are the biggest challenges facing the community in using patent data

	Not A Problem	Doesn't Really Affect My Work	A Significant Problem
Cleaning Patent Applicant Names	10	12	51
Obtaining Legal Status Information	21	9	43
Cleaning Inventor Names	13	19	41
Translating Patent Texts	20	18	36
Accessing Registries and file wrappers	18	20	29
Learning programming languages	19	25	26
Obtaining Family Data	31	15	25
Working with IPC codes	37	19	15
Obtaining Full Text	40	18	14
Working with CPC codes	37	20	14
Obtaining Patent Claims	42	16	13
Obtaining Images	34	24	11
Obtaining Sequence Data	21	37	9
Obtaining Patent Data (biblio front pages)	52	13	7

Ranked on column 3 to highlight significant problems

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

The open source tools era

Typical use of open source tools

- Typical user profile:
  - Researchers
  - Data scientists
  - IT
  - Business analysts

- Data typically analyzed through open source tools:
  - Non-structured data, structured non-patent data

  - Patent data: not a typical field where open source tools are being used

WIPO
WORLD
INTELLECTUAL PROPERTY
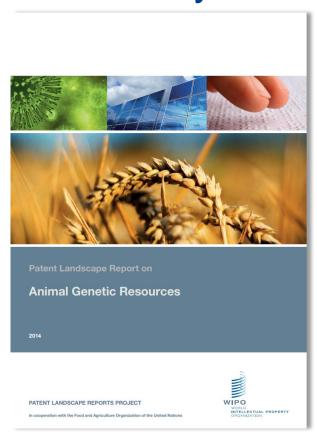ORGANIZATION

Pros and cons of open source tools

Advantages

- Free of charge
- A lot of them customizable to specific needs
- Give flexibility to experiment and choose the tool or combination of tools that is more adapted to the user´s needs
- Support and troubleshooting by programmers and other users in various fora

Disadvantages
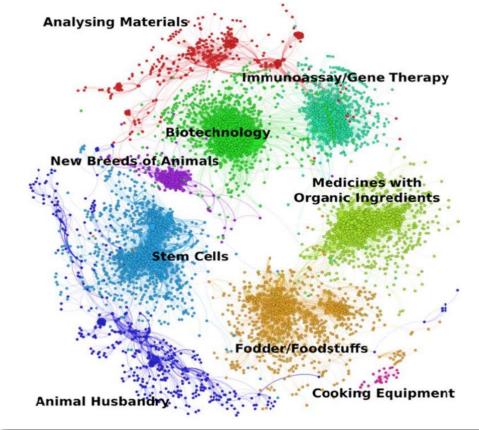
- Not as intuitive or user-friendly as commercial tools
- Require some programming skills, time and practice
- Several bugs/problems
- No commercial support

First thoughts on using open source tools for patent analytics

WIPO PLRs 2010-2014: exploring various commercial analysis and visualization tools

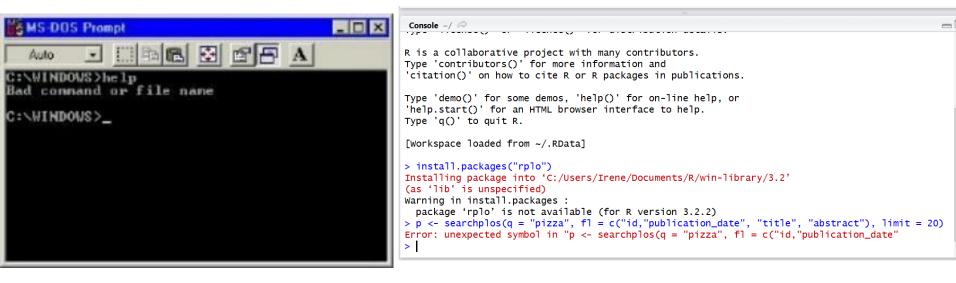2014: WIPO animal genetic resources PLR used for the first time Gephi, an open-source tool.

The Manual on Open Source Tools for Patent Analytics

- Project launched in May 2015, funded by FIT-JP
- Key author of the Manual: Paul Oldham
- Aimed at exploring:
  - various free and open source tools which could be used for various patent analysis tasks by users in developing countries
  - How to **obtain** data to use for patent analytics
  - How to **clean/tidy, analyze, visualize** and **share** patent data with the help of selected open source tools
- Challenge: programming skills required
- The Manual includes step-by-step instructions on the use of selected open source/free tools

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

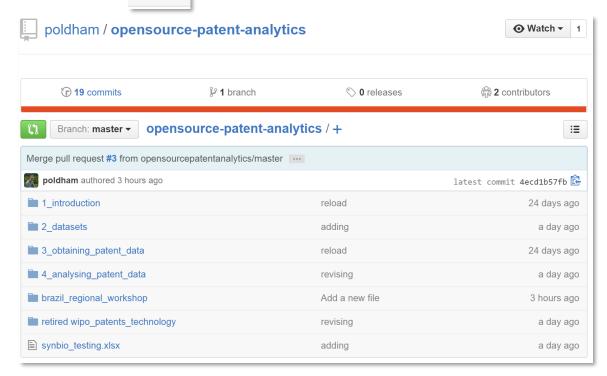# The process and the challenges

1987

2015

- ❑ Acquiring programming skills – sometimes steep learning curve

- ❑ Selecting the open source tools to explore and cover in the Manual

- ❑ Deal with constantly changing open source landscape

WIPO
WORLD
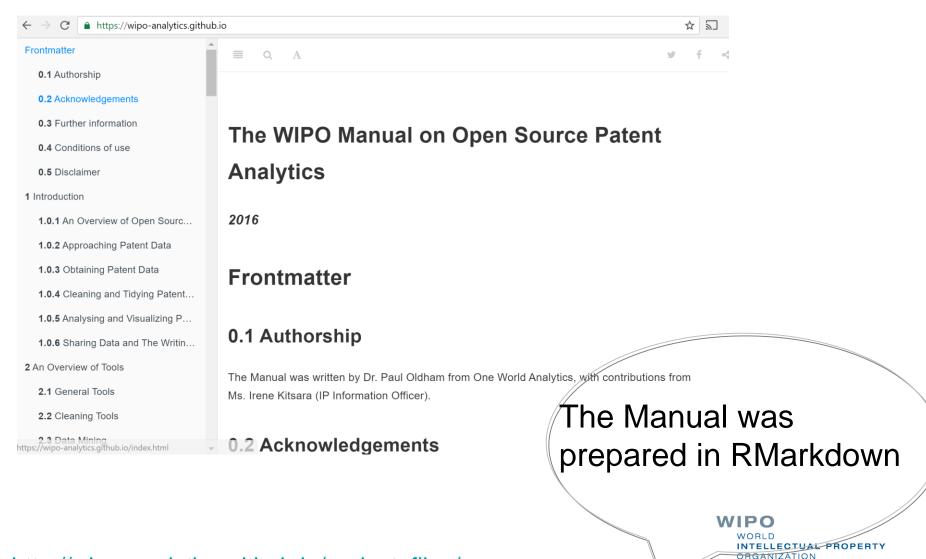INTELLECTUAL PROPERTY
ORGANIZATION

The project approach

A. The Project Development Site and its Repository – meet Github

Project development site with a possibility for comments and contributions by other parties

https://github.com/wipo-analytics/wipo-analytics.github.io

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

The Manual Repository with the Manual Chapters

The Manual was prepared in RMarkdown
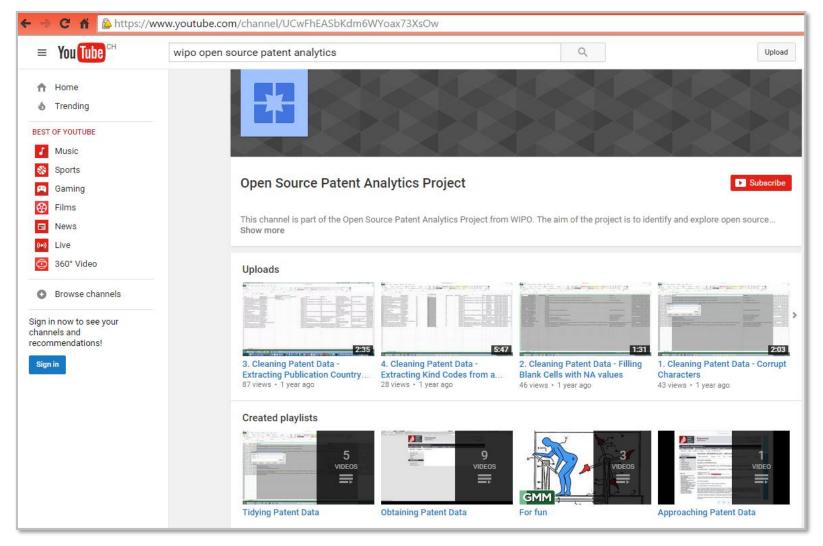
http://wipo-analytics.github.io/project_files/

The Manual Chapters

- Step-by-step instructions and guidance
- Screenshots
- Links to tutorials and further useful resources
- Tips and tricks

B. Video tutorials on individual analytics tasks

https://www.youtube.com/channel/UCwFhEASbKdm6WYoax73XsOw
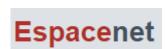
Which analytics tasks, databases and tools does the Manual cover?

PATENTSCOPE

- Obtaining Data

rplos

Patent 2 Net

Espacenet

OpenSci

This package is part of the rOpenSci project.
To learn more, visit rOpenSci.org

Lens

- Cleaning Data

R

R

- Visualizing Data

Gephi

tableau public

Plotly

ggplot2

- Sharing Data

GitHub

infogr.am

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

THANK YOU!