

数据使用手册

2023版

数据目录

1.中国全部专利申请与授权数据

- 提供1985年至最新年份的专利申请和专利授权数据，及专利事务数据，分发明公布、发明授权、实用新型和外观设计，详细追溯专利的各个事务状态；
- 含基本信息表、事务表；
- 字段：专利名称、申请公布日/号、授权公告日/号、申请人、发明人、地址、IPC分类号、PCT申请/公布日期、优先权、生物保藏、简要说明等。

2.中国专利引用数据（中国版、世界版）

- 提供1985年至最新年份所有中国专利在国内外的引用与被引用详细情况。世界版包含中国专利被本国申请的专利引用情况、世界上其他国家对中国专利的引用情况、该专利被引用情况的详细信息、该专利引用其他专利的详细信息以及各个专利在每个国家截止目前的法律状态信息；
- 含基本信息表、引用表、被引用表、事务表、文献引用信息表；
- 字段：专利申请公布/授权公告号、IPC分类号、引用及被引用专利名称等。

3.全球专利及引用被引用数据

- 囊括全球多个国家的发明专利、实用新型专利和外观设计专利基本信息及引用被引信息；
- 含专利基本信息表、专利引用信息表（外观设计无引用表）、专利被引信息表、专利事务表、专利IPC分类号表；
- 字段：专利申请公布/授权公告号、IPC分类号、引用及被引用专利名称等。

4.中国绿色专利申请与授权数据

- 根据最新的WIPO标准与最新IPC分类号划定，形成发明公布、发明授权、实用新型三类专利基础信息表及绿色专利类别表，结构清晰；
- 字段：专利名称、申请号、地址、授权公告号、绿色专利主(子)分类等。

5. 中国专利文本数据

- 大样本量文本类数据，时间区间为1985年至最新年份，主体为专利描述文本信息、专利权利要求文本信息、专利基本信息表；
- 字段：专利申请公布/授权公告号、IPC分类号、描述内容、声明内容等。

6.中国专利分地区统计数据

- 本数据依据最新行政区域代码，按照省份及城市进行地区划分，并以月份为最小统计单位，对全国专利申请和专利授权数据进行了细致统计；
- 字段：省份代码、省、城市编码、市、申请年份、申请月份、授权年份、授权月份、发明专利申请数量、实用新型申请数量、外观设计申请数量、发明专利授权数量、实用新型授权数量、外观设计授权数量。

专利与创新数据

数据目录

7.中国绿色专利分地区统计数据

- 本数据依据最新行政区域代码，按照省份及城市进行地区划分，并以月份为最小统计单位，对全国专利申请和专利授权数据进行了细致统计；
- 在分地区统计的基础上，以绿色专利的主分类和一级分类为统计标准分别进行处理，满足多维度的数据统计需求；
- 字段：省份代码、省、城市编码、市、申请年份、申请月份、授权年份、授权月份、绿色专利主分类、绿色专利一级分类、发明（实用新型）专利申请数量、发明（实用新型）专利授权数量等。

8.A股上市公司专利及引用被引用数据

- 参考科学的数据匹配与处理方法，将中国上市公司数据与中国全部专利申请与授权数据、中国专利引用数据进行了细致匹配。本数据库时间区间为1990年至最新年份；
- 分发明申请/发明授权/实用新型申请/实用新型授权/外观设计申请/外观设计授权六模块，并详细分为：基础信息表、引用表、被引用表、事务表（外观设计无引用表），根据上述详情表统计得出数量统计、质量统计数据表；
- 字段：公司名称、年度、申请公布日、授权公告号、引用及被引用专利申请公布号、申请号、IPC分类号等。

9.A股上市公司绿色专利及引用被引用数据

- 提供了以中国上市公司为主体的绿色专利及其相关基本信息数据、事务数据及引用被引用情况信息，不仅包括上市公司本身，还包含上市公司子公司、联营公司、合营公司的所有绿色专利信息；
- 分发明申请/发明授权/实用新型申请/实用新型授权四模块下详细分为：基础信息表、引用表、被引用表、事务表（外观设计无引用表，无绿色专利），根据上述详情表统计得出数量统计、质量统计数据表；
- 字段：公司名称、年度、申请公布日、授权公告号、引用及被引用专利申请公布号、申请号、IPC分类号等。

10.中国工业企业专利及引用被引用数据

- 参考科学的数据匹配与处理方法，将中国工业企业数据与中国专利申请与授权数据、中国引用数据细致匹配而成。工业企业数据截至2013年，专利数据截至最新年份，另新增文本数据；
- 发明申请/发明授权/实用新型/外观设计四类模块下详细分为：基础信息表、引用表、被引用表、事务表（外观设计无引用表），并统计得出数量统计、质量统计数据表；
- 字段：公司名称、年度、申请公布日、授权公告号、引用及被引用专利申请公布号、申请号、IPC分类号等。

专利与创新数据

数据目录

专利与创新数据

11.中国工业企业绿色专利及引用被引用数据

- 本数据库将工业企业专利引用被引数据与中国绿色专利申请与授权数据匹配，得出中国工业企业所属的绿色专利信息及引用被引情况；
- 发明申请/发明授权/实用新型三类表下详细分为：基础信息表、引用表、被引用表、事务表（外观设计无引用表，无绿色专利），根据上述详情表统计得出数量统计、质量统计数据表，另新增文本数据；
- 字段：公司名称、年度、申请公布日、授权公告号、引用及被引用专利申请公布号、申请号、IPC分类号、绿色专利号类别信息等。

12.中国高校专利统计数据

- 覆盖全国31个省份的2740所高校的专利信息；包含专利统计信息、专利引用信息两部分，划分申请与授权两个阶段，展示专利三种类型的统计情况；
- 字段：学校名称、曾用名、办学性质、所属城市、专利申请号、专利授权公告号、引用其他专利次数、IPC分类号、是否fmaily to family引用等。

13.中国高新技术企业名单数据

- 覆盖全国31个省、自治区、直辖市及5个计划单列市的高新技术企业；
- 字段：认定年份、批次、变更前后企业名称、更名时间、资质状态等。

14.中国高新技术企业专利统计数据

- 覆盖全国31个省、自治区、直辖市及5个计划单列市的高新技术企业的专利信息，包含专利统计信息、专利引用信息两部分；
- 字段：企业名称、专利申请号、专利授权公告号、引用其他专利次数、IPC分类号、是否fmaily to family引用等。

15.中国专利无效宣告数据

- 对专利无效决定公告信息进行汇总整理，并按照专利类型与中国专利申请于授权数据匹配，得出专利基本信息表；
- 字段：决定号、专利号、发明名称、专利权人、无效宣告请求人、发文日、专利申请公布号、申请号、地址等。

16.中国全部专利详细地址数据

- 将中国专利申请人的地址数据进行详细解析，统一细化至区县一级，并附带详细地址的经纬度信息；
- 字段：ID、邮政编码、地址、国家、省份、城市、行政区划代码、坐标等。

17.新三板公司专利及引用被引用数据

- 分为发明申请、发明授权、实用新型、外观设计四个模块，各模块包含基础信息表、引用表、被引用表、事务表四张表格；
- 字段：企业名称、申请号、申请公布日、授权公告号、引用及被引用专利申请公布号、IPC分类号、法律状态等。

数据目录

专利与创新数据

18.欧洲专利局专利数据

- 包含欧洲专利局专利的基本信息字段；
- 字段：专利名称、专利申请人（国别）、专利发明人（国别）、IPC分类号、CPC分类号、C-sets分类号、专利优先权、专利申请日、专利公开日、公开号集合等。

19.美国/英国/欧洲/日本/韩国专利及引用被引用数据

- 对英国/日本/欧洲/美国/韩国等国家的专利信息数据进行专项整理；
- 既包含了该专利引用其他专利的详细信息，又包含了该专利被其他引用情况的详细信息，同时还统计了各个专利截止目前的法律状态信息；
- 字段：专利ID、专利名称、申请公布号、申请公布日、专利申请号、授权公告号、发明人、专利权人 、分类号、引用其他专利次数、专利被引用次数、法律状态等字段。

20.美国/英国/欧洲/日本/韩国专利文本数据

- 对英国/日本/欧洲/美国/韩国等国家的专利文本数据进行专项整理；
- 本数据包含专利描述信息和专利权利要求信息。

21.税收调查企业专利及引用被引用数据

- 参考科学的数据匹配与处理方法，将中国工业企业数据与中国专利申请与授权数据、中国引用数据细致匹配而成。税调企业调查时间为2008-2015年，专利数据截至最新年份，另新增文本数据；
- 发明申请/发明授权/实用新型/外观设计四类模块下详细分为：基础信息表、引用表、被引用表、事务表（外观设计无引用表），并统计得出数量统计、质量统计数据表；
- 字段：纳税人识别号、申请年度、专利（独立/合作）申请/授权数量、（独立/合作）申请专利的技术复杂度、（独立/合作）专利公开1-5年累计被引用次数、专利名称、申请公布日、申请日、授权公告日、发明人、专利权人、专利代理机构、代理人、优先权、IPC分类号、简要说明等。

22.中国专利权人基本信息扩展数据

- 将向国家产权局申请专利的企业名单与中国工商注册企业基本信息数据匹配，得出中国专利权人基本信息扩展数据；
- 向国家产权局申请专利的企业名单包含：发明公布—申请人，发明授权/实用新型/外观设计—专利权人；
- 字段：专利申请/专利权人、企业名称、统一社会信用代码、工商注册号、组织机构编码、注册时间、注册地址、经纬度、行业分类等。

数据目录

中国工商注册数据

1.中国工商注册企业全信息

- 极具研究价值且覆盖面广泛的中国工商注册企业全信息，涵盖企业的基本信息、重点关注、知识产权、企业发展、经营状况五个方面，合计超过40张表格，数据量超2亿条，覆盖1950年至最新年份全国34个行政区的企业信息；
- 字段：社会信用代码、地址、企业名称、成立日期、法人、控股企业、变更记录、网站备案、专利信息、股权冻结、融资信息、行政许可等。

2.中国工商企业进入退出信息统计数据

- 按照年份、地区、行业、企业性质分类对工商注册企业基本信息数据进行统计处理，提供了二维、三维、四维的统计数据；
- 字段：会计年度、省份、城市、区县、行业门类、行业大类、企业性质、注册/注销/存续企业数量。

3.税收调查企业基本信息扩展数据

- 对参与全国税收调查的企业名单进行汇总整理，获取了工商注册基本信息；
- 字段：注册资本、注册年份、所在省市区、行业大类、经营范围等。

4.中国工业企业基本信息扩展数据

- 使用模糊匹配和精确匹配两种方式，将中国工业企业数据与工商注册企业基本信息数据进行匹配，补充了系列企业数据库在企业基本信息层面的缺失；
- 字段：公司名称、匹配公司名称、法定代表人、统一社会信用代码、工商注册号、组织机构代码、成立日期、注册地址等。

5.失信企业基本信息扩展数据

- 使用模糊匹配和精确匹配两种方式，将失信企业数据与工商注册企业基本信息数据进行匹配，补充了系列企业数据库在企业注册时间、组织机构代码、纳税人识别号、经营期限及地址等字段方面的缺失；
- 字段：公司名称、匹配公司名称、法定代表人、统一社会信用代码、工商注册号、组织机构代码、成立日期、注册地址等。

6.中国高新技术企业基本信息扩展数据

- 使用模糊匹配和精确匹配两种方式，将中国高新技术企业名单数据与工商注册企业基本信息数据进行匹配，补充了系列企业数据库在企业注册时间、组织机构代码、纳税人识别号、经营期限及地址等字段方面的缺失；
- 字段：公司名称、匹配公司名称、法定代表人、统一社会信用代码、工商注册号、组织机构代码、成立日期、注册地址等。

7.北上广深工商注册企业基本信息数据

- 囊括了北京、上海、广州和深圳四座城市企业的基本信息；
- 字段：注册时间、注册资本、企业性质、法定代表人、所属行业、所属区域。

数据目录

中国工商注册数据

8.各行业工商注册企业基本信息数据

- 通过2017年国民经济行业分类标准，采用两位行业编码（大类），从工商企业中提取各行业企业，并识别其工商注册基本信息；
- 行业包括：食品业、制造业、农业、采矿业、建筑业等；
- 字段：企业名称、公司类型、工商注册号、组织机构代码、行政区划、成立日期、注册资本、经营范围、城市编码、所在区域等。

9.各行业工商注册企业数量统计数据

- 以二维、三维、四维多个角度统计了1950年至今的工商注册情况，不仅统计了新增企业数量，还统计了注销企业数量，得出存续企业数量；
- 字段：会计年度、省市县、行业分类、注册/注销/存续企业数量等。

10.各省份/珠三角/长三角工商基本信息数据

- 在工商注册企业全信息数据中，提取各省份/珠三角/长三角辖区内范围内的企业样本，并识别其工商注册基本信息；
- 字段：企业名称、公司类型、工商注册号、组织机构代码、行政区划、成立日期、注册资本、经营范围、城市编码、所在区域等。

11.线上招聘公司基本信息拓展数据

- 采用精确匹配、模糊匹配两种方式，将线上招聘网站信息数据中的公司名单与工商注册企业基本信息数据匹配；
- 字段：企业名称、匹配企业名称、工商注册号、注册资本、注册时间、省市县、经纬度、行业分类、经营范围等。

12.核酸检测机构工商注册数据

- 依据全国核酸检测机构名单，获取了所有核酸检测机构的工商基础信息、股东信息、股东对外投资信息、工商变更信息等；
- 字段：企业名称、匹配企业名称、工商注册号、注册资本、注册时间、省市县、经纬度、行业分类、经营范围等。

13.专精特新“小巨人”企业工商注册基本信息数据

- 涵盖我国2019-2022年间合计4批专精特新“小巨人”企业的工商注册企业基本信息；
- 字段：企业名称、工商注册号、注册时间、注册地址、公司状态、行业分类名称及代码、企业简介、经营范围等。

14.中国专利权人基本信息扩展数据

- 见前【专利与创新数据】板块

15.中国企业商标/软著数据

- 涵盖我国工商注册企业持有的商标及软件著作权信息；
- 字段：企业名称、软著名称、发布日期、注册日期等。

数据目录

文本数据

1.中国各省份官方报纸数据
- 将四川、黑龙江、青海、广西、湖北、福建、上海、吉林、辽宁、内蒙古、贵州等各地区官方报纸中的报道详细信息进行了汇总整理；
- 字段：日期、版面、前标题、标题、副标题、图片、内容等。

2.A股上市公司股吧评论数据
- 以文本信息内容为主，充分展现投资情绪，直观展现各贴内容、发帖人账号注册省市、关注与被关注等多方面内容，各表之间以帖子ID和用户ID连接；
- 字段：帖子ID、股吧名称、帖子标题、帖子内容、回帖内容等。

3.地方政府留言板文本数据
- 全面获取2011年以来的各年份地方政府留言板问答的文本内容，内容丰富，且未来仍会持续更新；
- 使用者可根据留言类型、提问及回复时间、回复状态、回答组织等多种指标搜索，为公共管理、政府回应、治理绩效、等学术研究提供了很好的数据来源；
- 字段：话题、时间、回答组织、问题、回复内容、版块等。

4.中国专利文本数据
- 见前【专利与创新数据】板块

5.CCTV新闻联播文本数据
- 包含2006年中旬至今的文本类非结构化数据，字段简洁、内容丰富，并持续更新；
- 字段：发布日期、新闻标题及内容。

6.A股上市公司投资者关系管理数据
- 收集了上证e互动、深交所互动易、全景网的问答数据以及上市公司发布的投资者关系活动记录表中的文本信息；
- 字段：股票代码、提问内容/时间、回答内容/时间、提问者/回答者名称等。

7.A股上市公司全部公告数据
- 涵盖各季度及年报、公司治理、股东大会、日常经营等26个类型公告；
- 字段：股票代码、公司名称、公告标题、公告时间、分类、行业、年度等。

8.谣言数据-OPEN
- 较真查证平台上的辟谣信息；
- 字段：谣言ID、标题、内容、结论、解释源、作者、发布时间等。

9.唐诗宋词数据-OPEN
- 包含唐诗宋词文本信息、作者信息等；
- 字段：唐诗、宋词、宋诗作者/名/正文等。

10.A股财经报纸文本数据
- 见后【A股上市公司数据】板块

数据目录

11.中国各地区政府工作报告文本数据

- 涵盖全国29个省份及对应的304个地级市，时间跨度近20年，包含了国务院、省级政府、市级政府在内的三级政府部门的工作报告文本文件，多以txt格式文件展现，内容可用性极强；
- 字段：年份、层级、省份、城市、报告全文。

12.A股上市公司IPO申报发行文件数据

- 涵盖中国资本市场中各公司IPO上市流程中的各类文件；
- 字段：公司名称、披露类型、上市板块、保荐机构、披露时间、公告等。

13.中文金融情感词典-OPEN

- 中央财经大学姜富伟教授及其团队创建的公共词典数据；
- 字段：话题、时间、回答组织、问题、回复内容、版块等。
- 使用本数据需引用：姜富伟，孟令超，唐国豪. 媒体文本情绪与股票回报预测. 经济学(季刊),2021年第4期，第1323-1344页.

14.电影评论数据

- 全面涵盖了电影相关的所有评论，包含电影信息表、电影人员信息表、电影短评信息表、电影长评信息表、影评用户信息表，尤其包含大量丰富的长篇影评文本；
- 字段：电影ID、电影名称、导演、编剧、主演、类型、制片国家/地区、上映时间、片长、豆瓣评分、评分人数、评分分布、影评人ID、影评等。

15.中餐食谱数据-OPEN

- 包含十三万条中餐食谱信息；
- 字段：菜名、介绍、配料、标签、方式、时间、步骤等。

16.中国对联数据-OPEN

- 将对联数据进行了汇总整理，分为上联和下联字段；
- 字段：上联、下联。

17.CNN新闻文本数据

- 收集了22年来的CNN各专题节目的新闻文本内容，包括CNN Tonight、CNN This Morning、CNN Newsroom、Connect the World、At This Hour等多个新闻节目；
- 字段：新闻日期、所属媒体、新闻标题、新闻内容。

18.A股上市公司交易所监管措施数据

- 整理了上海证券交易所和深圳证券交易所公布的监管措施信息；
- 字段：证券代码、监管类型、处理事由、涉及对象、处理日期、文件名等。

19.A股上市公司新闻文本数据

- 见后【A股上市公司数据】板块

文本数据

数据目录

房地产与土地数据

1.中国小区数据

- 覆盖全国28个省级行政区、直辖市与对应91个重点城市的住房小区分布信息；更进一步地，本数据按城市类型进行划分，展现了一线、新一线、二线、三四线城市小区信息，针对性强，层次鲜明，具有极强的研究代表性；
- 字段：省份、城市、街道、小区名称开发商、建筑类型等。

2.中国二手房交易数据

- 覆盖全国28个省级行政区、直辖市与对应90多个重点城市的历史年份的二手房分布及交易信息，并按照房屋交易情况划分为二手房成交及挂牌信息数据；
- 字段：价格、地级市、区域、街道、交易权属、房屋年限、建筑面积等。

3.中国写字楼租赁数据

- 覆盖全国28个省市与对应多个重点城市的写字楼租赁信息；按照城市类型进行划分，展现了一线、新一线、二线、三四线城市的写字楼租赁情况；
- 字段：写字楼简介、租金、省市区、所在商圈、租赁方式、楼层、工位等。

4.中国写字楼买卖数据

- 覆盖全国28个省级行政区、直辖市与对应多个重点城市的写字楼买卖信息；按照城市类型进行划分，展现了一线、新一线、二线、三四线城市的写字楼买卖情况；
- 字段：写字楼简介、楼层、工位、朝向、位置、所在楼盘、楼盘价格等。

5.中国商铺租赁数据

- 覆盖全国28个省级行政区、直辖市与对应多个重点城市的商铺租赁信息；
- 字段：商铺简介、楼层、装修、位置、所在楼盘、商铺数量、商铺价格等。

6.中国商铺买卖数据

- 覆盖全国28个省级行政区、直辖市与对应多个重点城市的商铺买卖信息；
- 字段：商铺简介、楼层、朝向、装修、位置、楼盘名称、商铺价格等。

7.中国新房信息数据

- 覆盖全国28个省级行政区、直辖市与对应多个重点城市的新房信息；
- 字段：楼盘名称、项目地址、开盘时间、户型、价格、开发商等。

8.土地交易数据

- 涵盖土地出租、抵押、转让等全角度数据；
- 字段：行政区、电子监管号、支付期号、出租期限、土地抵押权人名称等。

9.法拍房数据

- 含房屋基本信息、竞买信息记录、拍卖公告表、标的物信息表、拍卖须知表；
- 字段：房屋名称、竞拍信息、处置单位、详细地址、变卖价、评估价等。

10.房地产业工商注册企业基本信息数据

- 见前【中国工商注册信息】板块

数据目录

A股上市公司数据

1.A股上市公司基本信息数据

- 包含基本信息表及股票曾用名表；
- 字段：公司名称、省份、实际控制人、主营业务、公司简介、高管人员、法定代表人、注册地址、办公地址、股票曾用名变更等。

2.A股上市公司收入结构数据

- 包含中国上市公司的主营业务介绍及所属行业信息、主要客户及供应商情况；
- 字段：证券代码、公司名称、主营业务、产品类型、经营范围、各行业/地区相关业务的营业收入及成本、主要客户及销售额、董事会经营评述等。

3.A股上市公司股票高频交易数据

- 包含国内上交所、深交所、科创板在内的股票市场逐笔成交订单情况；
- 字段：日期、交易流水号、成交时间、成交价、成交量、委买量、委卖量等。

4.A股上市公司及子公司数据

- 按年度对上市公司与其子公司、关联公司的关系与公司名称进行了整理；同时对上市公司母公司及其下属的关联公司进行深度整理，依据企业名称与国家工商企业数据进行匹配，采用精确匹配和模糊匹配两种方式分别进行处理，补充了上市公司的下属关联公司在很多基础信息方面的缺失；
- 字段：证券代码、会计年度、关联公司名称、关联公司与上市公司的关系、工商基本信息字段等。

5.A股上市公司股东信息数据

- 涵盖了中国上市公司各个报告期的股东结构信息、十大股东及十大流通股东的持股信息、基金持股信息；
- 字段：非/流通股非本、限售股本、变动原因、持股数量/比例等。

6.A股上市公司董监高信息数据

- 包含包含上上市公司历届高管、董事会和监事会人员数据；
- 字段：姓名、职务、起始/终止日期、名目、学历、国籍、持股数量等。

7.A股上市公司停复牌数据

- 涵盖了中国上市公司每日停复牌信息；
- 字段：证券代码、停复牌日期、停牌原因等。

8.A股上市公司业绩预告数据

- 涵盖了中国上市公司业绩预告及业绩快报数据；
- 字段：证券代码、公告日期、预告净利润上下限、预告摘要、变动原因、营业收入、营业利润、每股收益等。

9.A股上市公司全部公告数据

- 见前【文本数据】板块

数据目录

A股上市公司数据

10.A股上市公司财务报表数据

- 涵盖了中国上市公司各个报告期的财务信息，包括资产负债表、利润表、现金流量表、财务指标表、财报披露时间表；
- 字段：证券代码、报告期、货币资金、营业总收入、披露日期修正等。

11.A股上市公司重大事项数据

- 涵盖了中国上市公司在经营期间发布的重要事项，包括分红送股表、增发信息表、内部交易表、限售解禁表、股票回购表；
- 字段：证券代码、分红日期、每股分红、派息日、增发股数、增发价格、变动人、变动量、与董监高关系、回购数量/金额等。

12.A股上市公司机构调研数据

- 涵盖了中国上市公司历年来的机构调研信息；
- 字段：证券代码、公告日期、接待日期/方式/地点/人员、调研内容。

13.A股上市财务审计意见数据

- 涵盖了会计事务所对中国上市公司评估的财务审计结果；
- 字段：证券代码、公告日期、报告期、审计结果/费用、事务所、会计师等。

14.A股上市公司股东大会投票信息数据

- 收集整理了中国上市公司股东大会的召开情况，以及会议出席情况与决议情况，覆盖了中小股东在内的现场表决、网络表决比例；
- 字段：证券代码、公司名称、公告日期、会议类型、议题、审议内容、会议日期、股东数量、现场/网络投票股东数量、赞成/反对股份比例等。

15.A股上市公司融资融券数据

- 涵盖了中国上市公司在经营期间融资融券交易数据；
- 字段：证券代码、交易日期、收盘价、涨跌幅、买入额、偿还额、卖出量、偿还量、融资融券余额等。

16.A股上市公司专利及引用被引用数据

- 见前【专利与创新数据】板块

17.A股上市公司绿色专利及引用被引用数据

- 见前【专利与创新数据】板块

18.A股上市公司股吧评论数据

- 见前【文本数据】板块

19.A股上市公司新闻文本数据

- 涵盖来自近百家网络新闻媒体的近千万条中国上市公司新闻的全文文本数据，包含新闻来源及内容、公司名称等信息，为文本分析提供绝佳的研究来源；
- 字段：证券代码、新闻标题、新闻时间、新闻内容、报道媒体、原始媒体等。

数据目录

20.A股财经报纸文本数据

- 覆盖国内上海证券报、证券时报、第一财经日报等多个专业财经报纸文本信息，范围涉及证监会指定披露信息财经报纸及财经期刊、影响力大的政策导向财经报纸、影响力大的市场导向综合报纸、香港主流综合报纸等；
- 字段：日期、版面、标题、副标题、内容等。

21.A股上市公司投资者关系管理数据

- 见前【文本数据】板块

22.A股上市公司信息披露评分数据-OPEN

- 汇集了深交所与上交所对上市公司信息披露质量的打分情况；
- 字段：公司代码、公司简称、评价结果、年份。

23.A股上市公司证券交易所问询函数据-OPEN

- 涵盖上交所的问询函件、深交所的问询函件及回复；
- 字段：公司代码、公司简称、发函日期、监管问询类型、标题、文档名等。

24.A股上市公司招聘数据

- 包含中国上市公司近年的对外招聘信息；
- 字段：证券代码、公司名称、发布日期、职位、薪资、学历、工作经验、地区。

25.A股上市公司招投标数据

- 汇集了中国上市公司近年的招标及中标信息；
- 字段：证券代码、公司名称、公告日期、公告标题、项目名称、地区、招标类型、采购单位、采购单位地址、供应商、中标金额等。

26.证券从业人员信息数据

- 包含自2003年以来的证券从业机构人员信息和变更信息；
- 字段：姓名、性别、执业机构、登记编号、学历、登记状态等。

27.基金公司及从业人员信息数据

- 包含基金机构从业人员的基本信息和变更信息；
- 字段：姓名、性别、从业机构、证书编号、证书取得/变更日期、证书状态等。

28.A股上市公司IPO申报发行文件数据

- 见前【文本数据】板块

29.A股上市公司社会责任评价数据-OPEN

- 从股东责任、员工责任、供应商、客户和消费者权益责任、环境责任和社会责任五项考察，各项分别设立二级和三级指标对社会责任进行全面的评价；
- 字段：股票代码、股东/员工/环境/社会/责任、绩效、关爱员工、环境治理等。

30.A股上市企业数字化转型指数数据-OPEN

- 见后【公共数据】板块

数据目录

31.A股上市公司供应链数据

- 包含了上市公司主要客户表、上市公司主要供应商表；
- 字段：股票代码、截止日期、客户/供应商名称、销售/采购额、占比。

32.A股上市公司主营业务详情数据

- 包含了上市公司经营评述表、上市公司主营介绍表、上市公司主营构成表；
- 字段：股票代码、报告日期、报告内容、主营业务、所属行业、类别、产品名称、产品营收、收入占比等。

33.A股上市公司员工评价数据

- 包含公司实力情况表、公司评分情况表、员工评价内容表、薪资整体评价表、薪资岗位评价表；
- 字段：公司代码股票代码、1-5星评分占比（薪酬福利、工作条件、工作生活平衡、职业发展、公司认同）评论内容、提交时间、岗位名称、平均薪资等。

34.A股上市公司裁判文书数据

- 提供了不仅包括上市公司本身，还包含上市公司子公司、联营公司、合营公司的裁判文书信息；
- 字段：证券代码、关联公司名称、与上市公司关系、案件名称、金额、是否原/被告/第三方、案件成员、审判结果、裁决日期、裁判案号、案件原因等。

35.A股上市公司现场检查随机抽查数据

- 包含中国上市公司近年的现场检查随机抽查信息；
- 字段：省份和计划单列市、发文日期、工作开展时间、检查范围截止时间、所辖区域上市公司数、抽取时公司名称等。

36.A股上市公司财务报告审计意见数据

- 包含财务审计意见表和关键审计事项表；
- 字段：股票代码、公告日期、报告期、审计结果、审计总费用、会计事务所、签字会计师、关键审计事项分类、关键审计事项描述、事项应对方法等。

37.企业的经济政策不确定性感知指数-OPEN

- 见后【公共数据】板块

38.中诚信绿金ESG评级数据

- 见后【ESG数据】板块

39.投服中心行权数据

- 包含了上市公司行权情况表、上市公司行权问答表、上市公司行权新闻表；
- 字段：股票代码、是否行使、问题、回答、提问/回答时间、标题、内容等。

40.A股上市公司交易所监管措施数据

- 见前【文本数据】板块

A股上市公司数据

数据目录

公共数据

1.ESIEC中国企业创新创业调查数据-OPEN

- 中国企业创新创业调查（Enterprise Survey for Innovation and Entrepreneurship in China，简称ESIEC）由北京大学国家发展研究院、中国社会科学调查中心、共青团北京大学委员会及各高校相关组织联合开展的大型企业调查研究活动，是北京大学中国社会科学调查中心核心调查项目之一，由北京大学企业大数据研究中心组织实施。
- ESIEC旨在通过科学抽样和实地追踪调查，获得反映中国创新创业企业实际状况的微观数据，推动更高质量的学术和政策研究。调查内容主要包括企业家的创业史、企业创建过程、企业基本信息、企业创新、企业间关系以及营商环境等七个方面。
- 与北京大学企业大数据研究中心建立深度合作并建立数据索引。

2.中文金融情感词典-OPEN

- 姜富伟教授及其研究团队于2021年第4期《经济学（季刊）》发表了《媒体文本情绪与股票回报预测》，并在文中介绍此项金融学科研究成果。
- “我们使用该情感词典计算我国财经媒体文本情绪指标，发现媒体文本情绪可以更准确地衡量我国股市投资者情绪的变化，对我国股票回报有显著的样本内和样本外预测能力。媒体文本情绪对一些重要的宏观经济指标也有显著的预测能力，具有重要的学术和实践应用价值。”——《媒体文本情绪与股票回报预测》
- 经授权，建立了本数据的展示区及数据索引，便于浏览。

3. 1：100万全国基础地理数据-OPEN

- 全国1：100万基础地理数据覆盖全国陆地范围和包括台湾岛、海南岛、钓鱼岛、南海诸岛在内的主要岛屿及其临近海域，共77幅1:100万图幅，该数据整体现势性为2015年。数据采用2000国家大地坐标系，1985国家高程基准，经纬度坐标。
- 获得了国家基础地理信息中心的许可，将1:100万全国基础地理数据库收录于公共数据专区。

4. 1：25万全国基础地理数据-OPEN

- 全国1：25万公众版基础地理数据覆盖全国陆地范围和包括台湾岛、海南岛、钓鱼岛、南海诸岛在内的主要岛屿及其临近海域，共816幅。
- 获取到国家基础地理信息中心的许可，将1:25万全国基础地理数据收录于公共数据版块。

数据目录

公共数据

5. 30米全球地表覆盖数据-OPEN

- GlobeLand30数据是中国国家高技术研究发展计划(863计划)全球地表覆盖遥感制图与关键技术研究项目的重要成果。该数据集包含十个主要的地表覆盖类型，分别是耕地、森林、草地、灌木地、湿地、水体、苔原、人造地表、裸地、冰川和永久积雪。
- 获取到国家基础地理信息中心的许可，将30米全球地表覆盖数据收录于公共数据版块。

6.CHFS中国家庭金融调查数据

- 中国家庭金融调查（China Household Finance Survey, CHFS）是中国家庭金融调查与研究中心（以下简称“中心”）在全国范围内开展的抽样调查项目，旨在收集有关家庭金融微观层次的相关信息，主要内容包括：人口特征与就业、资产与负债、收入与消费、社会保障与保险以及主观态度等相关信息，对家庭经济、金融行为进行了全面细致刻画。
- CHFS基线调查始于2011年，目前已分别在2011、2013、2015、2017和2019年成功实施五轮全国范围内的抽样调查项目，2021年第六轮调查还在进行中。CHFS最新公开的2019年第五轮调查数据，样本覆盖全国29个省（自治区、直辖市），343个区县，1360个村（居）委会，样本规模达34643户，数据具有全国及省级代表性。
- 获取到中国家庭金融调查与研究中心的授权，将中国家庭金融调查数据收录于公共数据版块。

7.公司层面气候变化暴露指数-Firm-level Climate Change Exposure

- Contributors: Zacharias Sautner, Laurence van Lent, Grigory Vilkov, Ruishen Zhang
- Description: We introduce a method that identifies firm-level climate change exposures from conversation in earnings conference calls of more than 10,000 firms from 34 countries between 2002 and 2020.The exposure measures exhibit cross-sectional and time-series variations which align with reasonable priors, and are better in capturing firm-level variation than carbon intensities or ratings.The exposure measures relate to economic factors that prior work has identified as important correlates of climate change exposure (e.g., public climate attention).
- 经上海财经大学会计学院张瑞申教授授权，建立了该数据的展示区及数据索引，便于浏览。

数据目录

公共数据

8.A股上市公司社会责任评价数据-OPEN

- 见前【A股上市公司数据】板块

9.中国90座城市建筑物屋顶矢量数据集-OPEN

- 该数据集包含中国90座城市（根据城市行政等级及区域分布综合选取）建筑物屋顶矢量数据。主要基于深度学习语义分割模型和多源遥感影像进行制作。首先，对原始影像进行预处理，并根据城市等级及其区域分布情况进行分层采样以及目视解译，制作训练和测试数据。然后将训练数据输入深度学习语义分割模型进行训练，使其适用于建筑物屋顶提取任务，并基于测试数据，采用深度学习领域结果评价一般性指标对建筑物屋顶提取模型性能进行评价。最后，将此模型应用于中国90座城市建筑屋顶提取任务中，自动提取建筑物屋顶并进行矢量化。
- 获得了南师大智慧城市感知与模拟实验室的许可，将该数据收录于公共数据专区，以便浏览阅读 。

10.企业的经济政策不确定性感知指数-OPEN

- 2020年，聂辉华、阮睿和沈吉老师在《世界经济》杂志发表了《企业不确定性感知、投资决策和金融资产配置》一文，提供了一种计算企业层面的经济政策不确定性指数的方法，并利用该数据分析了不确定性感受对企业投资和融资的影响。
- 经授权，将本数据收录于公共数据版块，便于浏览。

11.中国省际铁路通行时间数据-OPEN

- 本数据来自南京大学长江产业经济研究院《全国统一大市场下的省际铁路交通研究报告》的附录部分。受铁路时刻众多、历史数据不容易搜集整理的限制，学术与政策研究者一直无法对铁路建设的时间节省效应进行准确估计。为此研究报告团队对历年《中国铁路旅客列车时刻表》的数据进行搜集，并通过对列车等级、铁路开行历史进行溯源分析，得到了2012-2022年间双数年份中国各省会城市（直辖市、自治区首府）之间的铁路列车通行时间，以期能从时间的角度反映中国铁路的建设成果。
- 获得了研究报告负责人的许可，将该数据收录于公共数据专区。

12.A股上市企业数字化转型指数数据-OPEN

- 广东金融学院吴非老师及其团队归集整理了上海交易所、深圳交易所全部A股上市企业的年度报告，并展开结构化分类（即“底层技术运用”与“技术实践应用”两个层面），形成特征词图谱，进而分类归集关键技术方向的词频并形成最终加总词频，从而形成了企业数字化转型指标统计基础数据库。
- 经授权，《金融经济学研究》（国家金融学专栏）与广东金融学院国家金融学学科团队共同决定，将企业数字化转型指数数据与研究报告免费对外界公开 。

数据目录

1.全球专利及引用被引用数据

- 见前【专利与创新数据】板块

2.美国/英国/欧洲/日本/韩国专利及引用被引用/文本数据

- 见前【专利与创新数据】板块

3.WIPO专利及引用被引用数据

- 囊括WIPO认证的发明、实用新型和外观设计专利的基本信息及引用、被引、法律状态信息；
- 字段：专利申请（授权）公布号、IPC分类号、引用及被引用专利名称等。

4.欧盟专利及引用被引用数据

- 囊括欧盟各国的发明、实用新型和外观设计专利的基本信息及引用、被引、法律状态信息；
- 字段：专利申请（授权）公布号、IPC分类号、引用及被引用专利名称等。

5.IFR工业机器人数据

- 囊括IFR工业机器人详细信息；
- 字段：年份、行业代码、行业、国家ISO代码、国家、新增机器人安装量等。

6.联合国大会投票信息数据

- 囊括两万多条自1946年以来的联合国大会会议投票信息，主表展示了各次会议的主要信息及各国投票明细情况；
- 字段：主题、标题、议程、决议、决定、投票、投票日期等。

7.机场信息数据

- 见后【地理与交通数据】板块

8.全球港口信息数据

- 见后【地理与交通数据】板块

9.全球地震信息数据

- 见后【生态与环境数据】板块

10.30米全球地表覆盖数据

- 见前【公共数据】板块

11.全球摩天大楼信息数据

- 见后【地理与交通数据】板块

12.电影评论数据

- 见前【公共数据】板块

13.国际足球比赛结果数据

- 见后【吃喝玩乐数据】板块

14.欧洲专利局专利数据

- 见前【专利与创新数据】板块

全球数据

数据目录

1.中国各省份官方报纸数据
- 见前【文本数据】板块

2.政府采购数据
- 覆盖2015以来的政府采购合同信息；
- 以2020年7月1日为界，分有2个版本，字段各有不同；
- 字段：合同金额、所属地域、供应方、采购方、所属行业等。

3.地方政府留言板文本数据
- 见前【文本数据】板块

4.中国各地区政府工作报告文本数据
- 见前【文本数据】板块

5.中国各区县最低工资数据
- 见后【中国分地区数据】板块

6.联合国大会投票信息数据
- 见前【全球数据】板块

7.CCTV新闻联播文本数据
- 见前【文本数据】板块

8.PPP项目数据
- 政企合作型城市基础设施项目数据，覆盖项目基本信息、准备阶段、采购阶段、执行阶段全数据；
- 各阶段有部分表格为材料文件内容，以文件名为主要显示方式。

9.失信企业基本信息扩展数据
- 见前【中国工商注册数据】板块

10.中国行政区划数据
- 见后【中国分地区数据】板块

11.国家标准数据
- 涵盖了所有中国现行、即将实施的和已经废除的推荐性、强制性标准；
- 字段：标准号、标准类型、标准状态、执行及主管单位、相近标准等。

12.国家基金项目数据
- 包含2019年度及以前的所有国家自然科学基金和社会科学基金项目，除立项项目的相关信息外，还涵盖详细项目负责人的信息、单位情况、成果情况等多方面的内容；
- 社科数据包含2019.07.15前所有数据，自科数据包含1987.01-2021.12期间所有数据；
- 字段：项目负责人、单位类别、结项时间、获奖情况、是否结题、对应学科及学部等。

政府数据

数据目录

政府数据

13.中国各地区政府招投标数据

- 涵盖了百余个招投标网站的招投标信息，总数据规模达4000万+，不仅覆盖了各地区的政府招投标平台，还包含了大量非官方平台的招投标信息，完整展现了招标、投标的过程情况；
- 字段：招标/中标单位、流程状态、行业类型、省市、预算金额、内容简介等。

14.中国彩票销售数据

- 对中国彩票销售数据进行系统整理，从全国、分类型、分地区等三方面汇总彩票销售情况；
- 有“中国福利彩票”和“中国体育彩票”两大系列，根据彩票特征又可分为乐透数字型、中福在线、即开型、视频型、基诺型、竞猜型等；
- 字段：年份及月度、福利/体育彩票销售额、当月销售额、累计销售额等。

15.中国游戏审批数据

- 从进口网络游戏、进口电子游戏、国产网络游戏等三方面汇总游戏审批情况，还提供了游戏审批变更及撤销信息；
- 字段：游戏名称、申报类别、出版/运营单位、文号、出版物号、时间等。

16.产品质量监督抽查数据

- 汇总国家监管总局公布的全国32个省市、2个特区的产品质量监督抽查数据；
- 字段：企业名称、所在省、产品名称、产品代码、产品登记、规格型号、生产日期/批号、抽查结果、承检机构、抽查时间/类型/单位、主要不合格项目等。

17.食品安全抽检数据

- 涵盖了我国自2014年以来的食品抽检结果信息，包括合格产品及不合格产品两个方面，同时也从食品层面和公司层面分开统计；
- 字段：包含食品名称、标称生产企业名称、抽检次数、不/合格次数、被抽检单位名称、所在省份、商标、标称生产企业地址、生产日期/批号、规格型号、抽检结果、检测项目名称、标准值、检测值等。

18.学科评估数据

- 涵盖了高校学科评估结果、科研单位学科评估结果；指标体系又具体分为“人文社科”、“理学”、“工学”、“农学”、“医学”、“管理学”、“艺术学”等七类；
- 学科评估平均四年开展一次，目前已开展4轮；
- 字段：轮数、大类名称、小类代码/名称、学校代码/名称、排名、评估结果、一级学科、单位代码/名称等。

19.境外投资企业（机构）名录数据

- 涵盖境外投资企业(机构)具体信息；
- 字段：境外投资企业（机构）、境内投资者名称、投资国别（地区）。

数据目录

政府数据

20.中国核酸检测机构及采样点数据
- 汇总整理的全国31个省份及直辖市的检测机构信息及核酸采样点信息；
- 字段：机构名称、采样点名称、地址、省份/地区、经纬度等。

21.中国企业税务行政处罚数据
- 整理了21世纪以来我国的税务行政处罚信息；
- 字段：企业名称、处罚日期/内容/事由、文件号等。

22.中国企业环保行政处罚数据
- 整理了21世纪以来我国的环保行政处罚信息；
- 字段：企业名称、处罚日期/内容/事由、文件号等。

23.无线电发射设备销售备案信息数据
- 整理了我国2019-2023年初的六万多条无线电发射设备销售企业备案信息；
- 字段：销售商名称、法人姓名、所在省市、详细地址等。

24.中国地方政府债券信息数据
- 整理了我国2009年以来的地方政府债券发行信息；
- 字段：债券名称、发行规模、发行日期、债券类型、利率等 。

中国分地区数据

1.中国各省份官方报纸数据
- 见前【文本数据】板块

2.中国工商企业进入退出信息统计数据
- 见前【中国工商注册数据】板块

3.中国各区县最低工资数据
- 包含各区县最低工资划分标准、金额和调整日期，展现了我国21世纪以来各县区层面的最低工资标准变化；
- 字段：省份、调整日期、划分标准、全/非全日制最低工资。

4.中国行政区划数据
- 分为三张子表：历年县以上的行政区划信息、历年县以下的行政区划变更信息；历年省市县乡镇街道及居委行政区划；
- 字段：年份、地区编码、省份、地级市、县区和县区级别、变更前后的乡镇名称和编码、变更原因、批准文件等。

5.中国（绿色）专利分地区统计数据
- 见前【中国专利与创新数据】板块

6.中国各地区方言信息数据
- 见后【历史文化数据】板块

数据目录

中国分地区数据

7.中国各地区碳交易数据
- 见后【ESG数据】板块

8.制造业工商注册企业数量统计数据
- 见前【中国工商注册数据】板块

9.农业工商注册企业数量统计数据
- 见前【中国工商注册数据】板块

10.北上广深工商注册企业基本信息数据
- 见前【中国工商注册数据】板块

11.农产品批发价格数据
- 见后【三农数据】板块

12.中国各地区政府招投标数据
- 见前【政府数据】板块

13.中国各地区政府工作报告文本数据
- 见前【文本数据】板块

14.中国彩票销售数据
- 见前【政府数据】板块

15.中国观鸟记录数据
- 见后【生态与环境数据】板块

16.全国兴趣点（POI）数据
- 共包含来自两个来源的POI信息，各来源在字段、数据量方面均有不同，其中，A来源包含15个大类数据；B来源包含17个大类数据；
- 分类一般有一级分类和二级分类，每个分类都有相应的行业的代码和名称一一对应；
- 基础信息：名称、别称、地址、坐标、电话、分类等；
- 详情信息：评分、菜单、价格、评论、团购/外卖信息、营业时间等；
- 衍生信息：分店数量、榜单排名等待挖掘产生的数据；
- 字段：商户名称、经纬度、省市县、详细地址、行业分类、各项评分（口味、环境、服务）、分店数量等。

17.中国核酸检测机构及采样点数据
- 见前【政府数据】板块

18.全国景区数据
- 汇总整理了全国31个省份及直辖市的1A-5A景区信息；
- 字段：景区名称、省份、景区级别、地址、经纬度、简介等。

数据目录

1.中国工业企业专利及引用被引用数据

- 见前【专利与创新数据】板块

2.中国工业企业绿色专利及引用被引用数据

- 见前【专利与创新数据】板块

3.中国工业企业股东信息数据

- 详细展现了1998—2013年规模以上工业企业的历史股东、最新股东以及股东信息变更的详细内容；
- 字段：企业名称、变更日期、变更项目、变更前、变更后、注册资本、企业法人、认缴/实际出资额、持股比例等。

4.中国工业企业主要人员信息数据

- 覆盖有1998—2013规模以上工业企业的董事长、总经理、执行董事、监事等涉及企业董监高的各类人员及其简介信息；
- 字段：企业名称、姓名、职务、简介、变更日期、变更前、变更后等。

5.中国工业企业信息变更数据

- 提供规模以上工业企业的工商历史变动信息，包含股东变更、主要人员变更、企业信息变更、经营变更、出资变更、其他变更等6个模块。
- 字段：企业名称、变更日期、变更项目、变更前、变更后。

6.中国工业企业基本信息扩展数据

- 见前【中国工商注册数据】板块

7.税收调查企业基本信息扩展数据

- 见前【中国工商注册数据】板块

8.税收调查企业专利及引用被引用数据

- 见前【专利与创新数据】板块

9.工业产品生产许可证数据

- 整理了全国31个省市工业产品生产许可证获证情况；
- 字段：公司名称、主导产品、证书编号、发证日期、许可证有效期至、说明、住所、生产地址、所在省份、明细等。

10.中国工业企业海关进出口注册数据

- 整理了我国中国工业企业海关进出口注册信息；
- 字段：企业名称、海关注册编码、注册日期、行业种类、注销标志等。

11.中国税收调查企业海关进出口注册数据

- 整理了我国中国税收调查企业海关进出口注册信息；
- 字段：企业名称、海关注册编码、注册日期、行业种类、注销标志等。

中国工业企业和税收调查企业数据

数据目录

1.中国绿色专利申请与授权数据
- 见前【专利与创新数据】板块

2.A股上市公司绿色专利及引用被引用数据
- 见前【专利与创新数据】板块

3.中国工业企业绿色专利及引用被引用数据
- 见前【专利与创新数据】板块

4.中国绿色专利分地区统计数据
- 见前【专利与创新数据】板块

5.中国各地区碳交易数据
- 覆盖国内八个碳排放交易所的碳排放交易数据，为碳中和研究助力；
- 字段：交易所、日期、成交量、成交价、成交额等。

6.公司层面气候变化暴露指数-OPEN
- 见前【公共数据】板块

7.A股上市公司社会责任评价数据-OPEN
- 见前【A股上市公司数据】板块

8.产品质量监督抽查数据
- 见前【政府数据】板块

9.食品安全抽检数据
- 见前【政府数据】板块

10.中国企业环保行政处罚数据
- 见前【政府数据】板块

11.工业产品生产许可证数据
- 见前【中国工业企业和税收调查企业数据】板块

12.中诚信绿金ESG评级数据
- 在评级依据的基础上，结合国内ESG信息披露现状和行业发展趋势，针对申万行业分类，中诚信绿金ESG评级方法共划分了57个行业评级模型，提取一级指标21项，二级指标55+项，三级指标180+项，全方位剖析ESG表现；
- 字段：证券代码、报告期、一/二/三级行业、E/S/G级别等。

13.电器电子产品有害物质声明数据
- 开发整理了电器电子产品有害物质声明数据；
- 字段：声明类型、认证机构名称、产品类别名称、证书状态、生产者（制造商）名称、产品名称等。

ESG数据

数据目录

三农数据

1.农业工商注册企业基本信息数据
- 见前【中国工商注册数据】板块

2.农业工商注册企业数量统计数据
- 见前【中国工商注册数据】板块

3.农产品批发价格数据
- 涵盖了全国31个省、直辖市，囊括45种主要农产品；
- 字段：交易所、日期、成交量、成交价、成交额等省份、农产品一/二/三级分类id、农产品一/二/三级分类名称、日期以及价格（元/公斤）。

4.中国种子进出口信息数据
- 涵盖了我国近十年的种子进出口审批信息；
- 字段：申请/出口单位、进出口国家、作物名称、供种单位、申请日期等。

5.中国种子储备数据
- 包含中国种子储备信息表、中国种子储备品种登记和审定信息表、中国种子储备品种权公告信息表、中国种子储备承储企业许可证信息表；
- 字段：作物名称、品种名称、申请号、申请日、申请人、申请状态、申请公告日、授权号等。

6.中国农药登记数据
- 涵盖了1986年以来的农药登记信息和近两千家已登记的企业信息；
- 字段：登记证号、农药名称、农药类别、剂型、总含量、有效期至、总有效成分含量、防治对象、用药量等。

7.中国肥料登记与备案数据
- 包含部级登记产品信息表、部级产品备案信息表、境外产品备案信息表与省级产品备案信息表；
- 字段：企业名称、产品名称、产品形态、技术指标、备案号、备案时间等。

地理与交通数据

1.机场信息数据-OPEN
- 含世界机场数据、中国机场数据（2020.01.15之前）；
- 字段：机场名、经纬度、机场地址、电话、开通年度月度。

2.铁路信息数据
- 含12306中所有列车时刻信息（2007， 2011， 2013及之后每年）、所有高铁线路表及各高铁站点开通时间表；
- 字段：列车编号/名称、起始及终点站编号/名称、出发/到达时间、途径站名称、列车类型、运行时间等。

数据目录

地理与交通数据

3.全球港口信息数据
- 包含全球两百多个国家及对应四千多个港口的基本信息；
- 字段：中英文名称、代码、所属航线、所属国家、简介等。

4.公交线路信息数据
- 包含全国200多个城市的公交线路信息；
- 字段：起/终点首/末班车时间、全程票价、票制、线路长度等。

5.飞机航班信息数据
- 包含国内机场航班详细信息；
- 字段：起飞机场、到达机场、起飞城市、到达城市、航班号、航班公司等。

6.交通运输、仓储和邮政业工商注册企业基本信息数据
- 见前【中国工商注册数据】板块

7.全国快递站点数据
- 涵盖了全国34个省、直辖市、自治区的快递网络，包含京东物流、顺丰速运、中通快递、申通快递在内的35家快递公司网点信息，并识别了各网点的经纬度信息；
- 字段：网点类型/名称、省市县、网点地址、（不）派送区域、经纬度等。

8.人口迁徙大数据
- 展现了2019年以来我国各城市春运及国庆期间的人口迁入及迁出情况，含各地区迁入来源地信息、各地区迁出目的地信息和各城市内部出行强度信息三个子模块，涵盖迁入和迁出城市、省份的总趋势和迁入迁出明细数据；
- 字段：迁入/出城市、迁徙规模指数、迁入/出人口占比、城内出行强度等。

9.全国兴趣点（POI）数据
- 见前【中国分地区数据】板块

10.1:100万全国基础地理数据
- 见前【公共数据】板块

11.1:25万全国基础地理数据
- 见前【公共数据】板块

12.30米全球地表覆盖数据
- 见前【公共数据】板块

13.全球摩天大楼信息数据
- 覆盖中国乃至全球多个国家摩天大楼的详细数据；
- 字段：高度、层数、建筑官方名称、建筑类型、建造状态、所在国家、所在城市、地址、邮政编码、功能、建筑全称、官网、其他名称、预计完工时间、结构材料、完工时间、能级、当前功能、原始功能等。

数据目录

地理与交通数据

14.全国景区数据
- 见前【中国分地区数据】板块

15.中国船舶用产品供应商名录
- 包含截至2021年的中国船舶用产品供应商的详细信息；
- 字段：证书编号、有效期、产品名称、工厂名称、公司名称、联系人地址等。

16.中国在册船舶信息数据
- 根据CSA和CSAD分类，识别国际与国内1975年至今在中国注册的船舶信息；
- 全面披露了全球三十多个国家在中国注册的船舶及其详细信息
- 字段：船名、船舶归属类别、船东、船旗国、船籍港、船区、航速、平均吃水、货仓数等。

17.中国省际铁路通行时间数据-OPEN
- 见前【公共数据】板块

18.中国城际轨道交通数据
- 涵盖我国各城市的地铁线路及票价信息；
- 字段：省份、城市、起/终点、线路、票价等。

人物数据

1.A股上市公司董监高信息数据
- 见前【A股上市公司数据】板块

2.证券从业人员信息数据
- 见前【A股上市公司数据】板块

3.基金公司及从业人员信息数据
- 见前【A股上市公司数据】板块

4.中国工业企业股东信息数据
- 见前【中国工业企业和税收调查企业数据】板块

5.中国工业企业主要人员信息数据
- 见前【中国工业企业和税收调查企业数据】板块

6.注册审计师信息数据
- 涵盖经中国注册会计师协会注册的会计师事务所及会计师的基本信息；
- 字段：事务所名称/编号、办公地址、批准设立机关、注册会计师人数、从业人员人数、小于等于40岁人数、本科人数、被检查信息、境外分支机构、姓名、性别、职务、学历、专业、资格取得方式等。

7.中国记者信息数据
- 涵盖中国记者有效身份证件（即新闻记者证）核发和注销信息；
- 字段：姓名、单位、核发/注销日期、公告日期、记者证号。

数据目录

海关与国际贸易数据

1.中国海关统计数据

- 涵盖了来自中华人民共和国海关总署的进出口货物贸易统计数据；
- 含进出口商品年度总值、月度总值、国别、商品构成、商品类章总值、贸易方式总值、企业性质总值表、收发货人所在地总值、目的地/货源地总值、商品关别总值、特定地区进出口总值、外商投资企业进出口总值等。

2.中国对外贸易指数数据

- 涵盖了来自中华人民共和国海关总署公布的中国对外贸易指数数据；
- 字段：HS编码、SITC编码、BEC编码、日期、参照基准、商品名称、数量/价格/价值指数等。

3.中外双边贸易数据

- 涵盖了中华人民共和国商务部的国别贸易数据，包括了亚洲、欧洲、非洲、美洲、大洋洲等在内5大洲共计59个国家及中国香港、中国台湾等2个地区的数据；
- 字段：国别、年月、HS编码、商品类型、本年金额、进出口金额、占比等。

4.国际货物贸易数据

- 涵盖了中华人民共和国商务部的国别货物贸易数据；
- 字段：国别、年月、进出口国家、进出口金额/同比/占比、HS编码等。

5.境外投资企业（机构）名录数据

- 见前【政府数据】板块

6.国际货运代理信息数据

- 涵盖了七万多家企业从事国际货运代理业务的相关信息，同时将企业名单与中国工商注册企业基本信息数据匹配，整理出国际货运代理企业信息增强数据；
- 字段：企业名称、经营场所、企业类型、运输方式、货物类型等。

7.外贸供应商名录数据

- 根据国内的B2B平台整理的外贸供应商名录数据；
- 字段：企业名称、地址、主营产品、标签、经营模式、商品数量、年交易额等。

生态与环境数据

1.台风数据

- 涵盖了我国东南沿海至西太平洋的台风相关信息；
- 包含2个来源数据，各来源在字段、数据量方面均有不同，A来源包含台风基本信息、台风路径信息、台风风圈半径三张子表；B来源包含台风基本信息表、台风登陆信息表、台风路径信息表、台风预测信息表四张子表；
- 字段：台风编号、中英名称、发生日期、台风强度、经纬度、风速、移动方向、路径编号、风速级别等。

数据目录

生态与环境数据

2.空气质量站点监测数据-OPEN

- 各城市空气质量的日度数据、位置信息、监测数据（自2014年5月13日起）；
- 字段：监测点编码、名称、城市、经纬度、数据采集日期、空气质量指数等。

3.中国历史天气数据-OPEN

- 中国所有城市（含县）自2011年以来的历史天气数据，指标包含温度、天气、风向风力、空气质量（自2016起）等内容；
- 字段：省份、区县、日期、最低最高温度、风向、空气质量指数。

4.全球地震信息数据

- 全面、完整地涵盖了中国地震台网和美国地质调查局汇总的地震相关信息，指标丰富，内容全面；
- 含地震基本信息、周边城市、历史地震、震相记录、震测台网分布信息表；
- 字段：年份、发震时刻、经度、纬度、深度、震级、参考位置、周边城市、震相记录数等。

5.水利、环境和公共设施管理业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块

6.公司层面气候变化暴露指数-OPEN

- 见前【公共数据】板块

7.中国观鸟记录数据

- 涵盖了中国自1997年以来的中国观鸟记录信息，囊括34个省、直辖市、自治区，生动立体地展示了我国二十一世纪以来的观鸟发展历程；
- 字段：报告编号、观测时间/地点、坐标定位、记录用户、观鸟数量、鸟种编号、鸟中英文名及拉丁名、次数等。

8.1:100万全国基础地理数据库-OPEN

- 见前【公共数据】板块

9.1:25万全国基础地理数据-OPEN

- 见前【公共数据】板块

10.30米全球地表覆盖数据-OPEN

- 见前【公共数据】板块

11.全国景区数据

- 见前【中国分地区数据】板块

12.电器电子产品有害物质声明数据

- 见前【ESG数据】板块

数据目录

银行、保险与债券数据

1.中国银行网点全集数据
- 包含中国所有银行网点的分布信息，以银行网点持有的金融许可证状态为线索，分为金融许可证持有表、机构失控表、机构退出表；
- 字段：机构所在分局、机构名称、机构所在地、发证机关、发证日期等。

2.中国银行分支机构违规被处罚数据
- 除涵盖2004年以来中国银保监会官网公布的行政处罚决定书文中全部内容外，还整合了被处罚银行的经纬度，便于定位和进行空间计量分析；
- 字段：被处罚单位、法人、案由、依据、处罚决定、做出处罚的机关名称、处罚时间、被处罚银行的经纬度。

3.上海票据交易所票据承兑信息数据
- 搜集整理了商票信息披露平台披露的票据承兑信息详细数据，分为企业票据承兑信息情况表和财务公司票据承兑信息情况表；
- 字段：承兑人名称、统一社会信用代码、机构类别、注册日期、披露信息时点日期、承兑人开户机构名称、累计承兑发生额（元）、披露日期等。

4.中国邮政储蓄银行网点信息数据
- 包含全国31个省、直辖市及对应300多个地级市逾5万个网点分布信息、营业状态信息及营休情况；
- 字段：省市区、服务网点名称、邮编、地址、营休时间、营业状态等。

5.金融业工商注册企业基本信息数据
- 见前【中国工商注册数据】板块

6.中国保险机构网点全集数据
- 涵盖了保险机构网点分布、失控、退出3方面的信息；
- 字段：机构名称、机构住所、发证/批准/失控/退出日期、发证机关等。

7.中国保险中介机构网点全集数据-OPEN
- 涵盖了中国保险中介机构网点分布、失控、退出3方面的信息；
- 字段：机构名称、机构住所、发证/批准/失控/退出日期、发证机关等。

8.中国理财产品数据
- 涵盖了来自银行和保险公司发布的理财产品信息数据；
- 字段：产品名称、保险类别、投保年龄、保障期限、发行银行、币种、发行日、停售日、管理期、起始金额、收益类型、预期收益率等。

9.中国地方政府债券信息数据
- 见前【政府数据】板块

10.债券信息数据
- 涵盖债券信息和发行主体信息两个部分；
- 字段：债券代码、发行总额、利率、发行时面值、发行时债项/主体评级等。

数据目录

新冠防疫数据

1.核酸检测机构工商注册数据

- 见前【中国工商注册数据】板块

2.新冠病毒防疫专利信息数据-OPEN

- 涵盖中国及世界范围的新冠肺炎治疗用药、预防用药、检测和诊断试剂盒、医疗器械、防护产品、医用消毒、医疗废弃物处理、废水处理等多个领域；
- 字段：专利申请日、分类号、法律状态、申请号等。

3.新冠病毒防疫企业捐赠数据-OPEN

- 涉及到新冠病毒防疫的企业捐赠数据进行整理；
- 字段：捐赠人、募捐机构、疫情捐赠额、日期等。

4.新冠病毒防疫标准信息数据-OPEN

- 收录了全球范围内与新型冠状病毒肺炎防治相关的医用防护产品、传染病检验检测、医用基础设施等标准，为抗击疫情工作提供标准信息支撑；
- 字段：产品类别、实施日期、发布单位、设施类别、标准编号等。

5.全国小区新冠病毒确诊病例数数据

- 该数据库共计11868行数据，可从时间与空间两个维度探寻疫情发展趋势；
- 字段：详细地址、经纬度、及其所属省、市、县区、确诊人数和更新时间。

6.中国核酸检测机构及采样点数据

- 见前【政府数据】板块

美国数据

1.美国国家民主基金会捐赠数据

- 涵盖来源于美国国会的年度财政拨款的捐助项目情况，主要媒介渠道则是美国新闻署；
- 字段：项目名称、项目地区、项目国家、项目目的、受捐赠组织名称、捐赠数额、项目描述等。

2.美国专利及引用被引用数据

- 见前【专利与创新数据】板块

3.Glassdoor美国公司员工评价数据

- 美国公司员工对雇主的评价（包括对工作、公司、环境的满意度），以及薪酬福利、文化价值观、企业展望、是否推荐等多方面指标；
- 字段：公司名称、总部地址、公司行业、员工评价、回复日期、评论数、评分、是否在职等。

4.CNN新闻文本数据

- 见前【文本数据】

数据目录

教育、劳动与就业数据

1.线上招聘网站信息数据
- 涵盖了多个线上招聘网站信息，拥有跨时7年以上的历史日度数据；
- 字段：公司名称、地址、行业、岗位、职位描述、福利待遇、学历等。

2.中国高等学校名单数据-OPEN
- 涵盖了教育部公布的我国高等学校信息；
- 字段：学校名称、学校标识码、主管部门、所在地、办学层次、备注等。

3.中国各区县最低工资数据
- 见前【中国分地区数据】板块

4.教育业工商注册企业基本信息数据
- 见前【中国工商注册数据】板块

5.国家基金项目数据
- 见前【政府数据】板块

6.学科评估数据
- 见前【政府数据】板块

7.A股上市公司招聘数据
- 见前【A股上市公司数据】板块

8.Glassdoor美国公司员工评价数据
- 见前【美国数据】板块

9.A股上市公司员工评价数据
- 见前【A股上市公司数据】板块

10.线上招聘公司基本信息拓展数据
- 见前【中国工商注册数据】板块

年鉴数据

1.中国环境统计年鉴数据
- 收录了历年全国各省、自治区、直辖市环境各领域的基本数据和主要年份的全国主要环境统计数据；
- 对中国历年环境年鉴进行了系统性的整理，涵盖了自然状况、水环境、海洋环境、大气环境、固体废物、自然生物、土地利用、林业、自然灾害及突发事件、环境投资、城市环境、农村环境等十二个方面的环境统计资料。

2.中国审计年鉴数据
- 对中国历年审计年鉴进行了系统性的整理；
- 涵盖了全国审计机关、审计署派出审计局、审计署驻地方特派员办事处、地方审计机关这四个方面的综合审计情况。

数据目录

历史文化数据

1.中国宗教场所地理分布数据

- 展现中国佛教、道教的宗教活动场所信息；
- 字段：宗教、派别、场所名称、地址、省市区、经纬度等。

2.中国各地区方言数据

- 根据《中国语言地图集》，整理出1986版与2012版；
- 字段：省市县、经纬度、方言大类、分（小）片区等。

3.全国文物商店及拍卖企业数据-OPEN

- 涵盖现有的国有文物商店，经营文物的拍卖企业、非国有文物商店信息；
- 字段：注册地、主题类型、省内分类序号、名称、备注。

4.不可移动文物数据

- 涵盖了中国历史文化名城、名镇、名村、街区名单及中国世界文化遗产名录，还汇集了全国重点文物保护单位名单、革命文物保护利用片区分县名单；
- 字段：名城/镇/村/文化街区名称、省份、批次、文物名称、分类号、文物类型、时代、地址等。

5.博物馆和馆藏文物数据

- 整理收录了全国31个省市博物馆名录信息及馆内珍贵文物数据；
- 字段：省份、博物馆名称、性质、质量等级、是否免费开放、藏品数、珍贵文物数、教育活动数、展览数、参观人数、普查登记号、藏品名称、藏品类别、藏品年代、收藏单位 等。

6.中国被盗（丢失）文物数据-OPEN

- 包含中国被盗（丢失）文物信息发布平台披露信息；
- 字段：发布编号、类型、文物名称、年代、质量、完残情况、特征等。

7.中国文物保护工程数据

- 含文物保护工程资质单位信息表、文物保护工程资格人员信息表2张表；
- 字段：工程资质单位名称、经济性质、省份、资质类型、地址、资质公布文件、工程资格人员姓名、所在单位、资格类型、业务范围、证书编号等。

8.中华老字号企业名录-OPEN

- 整理收录了一千多家的中华老字号企业信息；
- 字段：企业名称、代表性注册商标、省份、单位地址等。

9.文化、体育和娱乐业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块

10.儒家文化数据

- 涵盖了中国孔庙信息、明朝书院信息、明朝学院信息及明清贞节牌坊信息；
- 字段：旧址、朝代、现属地省市县（代码）、书院/学校名称、姓名、事迹等。

数据目录

历史文化数据

11.中国博物馆数据

- 包含我国历年博物馆基本报告信息；
- 字段：博物馆名称、性质、等级、年参观量、地址、经纬度等。

12.中国家谱数据

- 覆盖大量人物、文化、民俗等资料；
- 字段：姓氏、撰修时间、堂号、家谱简介、馆藏机构、先祖名人等。

医疗、卫生与食品数据

1.医疗信息大数据

- 包含医院信息、科室信息和医生信息三张子表；
- 字段：医院详细位置、医院等级、科室名称、所属医院、医生人数、医生个人简历、主治领域、医师级别等。

2.中国药品上市与在审信息数据

- 包含药品受理品种表、临床试验默示许可表、上市药品表、原辅包登记信息表等，生动展现了我国历年药品在审、上市的一系列过程信息；
- 字段：药品名称、药品类型、企业名称、评审状态、受理号、承办日期、注册分类、企业地址、品种名称等。

3.中国医保服务信息数据

- 覆盖我国医保服务下涉及的定点医疗机构信息、定点零售药店信息、医保药品及医用耗材生产情况及异地联网定点机构信息，并提供了各机构的经纬度信息。
- 字段：医疗机构/药店名称、医保结算等级、详细地址、经纬度、药品名称、药品企业、批准文号等。

4.化学药品目录集

- 包含我国经评估认定纳入的目录集中的化学药品详细信息，并附有大量的药品使用说明书文件，是药品相关研究重要的文本数据；
- 字段：批准文号、药品名称、剂型、上市许可持有人、生产厂商、首次批准日期、上市销售状态、收录类别等。

5.卫生和社会工作业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块

6.食品业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块

7.食品安全抽检数据

- 见前【政府数据】板块

数据目录

8.餐饮外卖商家数据

- 包含不同来源的北京、上海、西安、重庆等25个城市信息，不同城市的不同来源在数据量和截至时间方面各不相同；
- 字段：商家名称、地址、所在省市县、经纬度、总评分、月售、配送费（元）、配送满意度等。

吃喝玩乐数据

1.中国电影票房数据

- 含电影基本信息表及票房详细信息表（2019.10.15前）；
- 字段：电影名、上映年度、类型、累计票房、导演及主演、单周及累计票房等。

2.电影评分数据

- 见前【文本数据】板块

3.国际足球比赛结果数据-OPEN

- 1872年至2017年间的足球赛事结果数据；
- 字段：比赛日期、主队及客队名字、比赛名称、全队得分（不包括点球）。

4.中餐食谱数据-OPEN

- 见前【文本数据】板块

5.中国对联数据-OPEN

- 见前【文本数据】板块

6.美国金球奖获奖数据-OPEN

- 将历年美国金球奖获奖信息进行了整理汇总；
- 字段：电影年份、颁奖年份、届数、奖项名称、被提名人、电影名称等。

7.中国彩票销售数据

- 见前【政府数据】板块

8.中国游戏审批数据

- 见前【政府数据】板块

9.中国星级酒店数据

- 收录了全国32个省市的一至五星级酒店名录；
- 字段：省份、星级、酒店名称、地址、城市等。

10.住宿和餐饮业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块

11.全国景区数据

- 见前【中国分地区数据】板块

12.文化、体育和娱乐业工商注册企业基本信息数据

- 见前【中国工商注册数据】板块