



中国开放数据 使用手册

2022版

CnOpenData 平台简介

Introduction

CnOpenData是覆盖经济、法律、医疗、人文等多个学科维度的综合型数据平台，并持续提供个性化数据定制服务，现已拥有70+个专题数据库，涵盖专利数据（1.4亿+量级）、工商注册企业数据（1.7亿+量级）、上市公司数据、工业企业数据、房地产与土地数据、政府数据、环境数据、分地区数据、交通数据、气象数据等多个数据系列，数值型数据和文本型数据并存，学术研究价值极高。

本平台集灵活的数据定制服务、专业度极强的数据整合能力、强大的学术性于一体，为各领域不同主体的研究决策提供优质、专业的数据支持。我们不仅为国内多所高校和著名学者提供持续的数据服务，在海外我们也入选了普林斯顿图书馆“来自中国的数据”系列。同时，我们与Neudata达成战略合作伙伴关系，也致力于为投资管理、私募股权、风险投资和企业领域的客户提供高质量的数据服务。

简约、清晰、有用，是我们追求的理念。

数据是散落的金子，我们愿做挖矿者，期待与作为炼金人的您精诚合作！

CnOpenData团队

2022-3-2



1. 中国全部专利申请与授权数据

- 提供1985-2021年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1985-2021年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WIPO标准划定七类绿色专利，并各自形成发明公布、发明授权、实用新型三类，结构清晰；
- 含发明申请表、发明授权表、实用新型表及对应所属绿色专利类别表；
- **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1985-2021年，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：**专利名称、专利申请公布/授权公告号、IPC分类号、描述内容。

6. 中国专利分地区统计数据

- 本数据依据2020年行政区域代码，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。

7.中国绿色专利分地区统计数据

- 本数据依据2020年行政区域代码，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.中国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1990-2021年；
- 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.中国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2018年；
- 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。

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.中国高等学校名单数据-OPEN

- 整理了教育部公布的我国高等学校名单，分普通高等学校和成人高等学校；
- **字段：**学校名称、学校标识码、主管部门、所在地、办学层次、备注。

13.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，统计数据以申请与授权两个阶段划分，展示专利三种类型的统计情况。专利引用数据分基础信息表、引用表、被引用表、事务表三张表格，分发明申请、发明授权、实用新型、外观设计四模块；
- **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family to family引用等。

14.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。

15.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，统计数据以申请与授权两个阶段划分，展示专利三种类型的统计情况。专利引用数据分基础信息表、引用表、被引用表、事务表三张表格，分发明申请、发明授权、实用新型、外观设计四模块；
- **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family to family引用等。

16.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；
- 含专利无效宣告信息表、无效宣告专利详情表；
- **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。

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1.7亿条，覆盖1950年至2021年全国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.中国工商注册企业数量统计数据

- 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、企业性质、注册/注销/存续企业数量。

3.中国上市公司基本信息数据增强版

- 使用模糊匹配和精确匹配两种方式，将中国上市公司名称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。

4.中国工业企业基本信息数据增强版

- 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。

5.失信企业基本信息数据增强版

- 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数据增强版

- 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. 食品行业工商基本信息数据

- 通过国民经济行业分类标准，采用两位行业编码（门类），从工商企业中提取食品行业企业，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。

8. 食品行业食品安全数据

- 涵盖国家食品药品监督管理局对食品行业企业进行的食品抽检信息；
- **字段：**企业名称、抽检详情、抽检结果、抽查产品、通报时间、通报文号等。

9. 零售业工商基本信息数据

- 通过国民经济行业分类标准，采用两位行业编码（门类），从工商企业中提取零售业企业，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。

10. 房地产业工商基本信息数据

- 通过国民经济行业分类标准，采用两位行业编码（门类），从工商企业中提取房地产业企业，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. 汽车制造业工商基本信息数据

- 通过国民经济行业分类标准，采用两位行业编码（门类），从工商企业中提取汽车制造业企业，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。

12. 浙江省工商基本信息数据

- 在工商注册企业全信息数据中，提取浙江省范围内的企业样本，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。

13. 珠三角工商基本信息数据

- 在工商注册企业全信息数据中，提取珠三角辖区内的企业样本，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。

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多个重点城市的二手房分布信息，并按照房屋交易情况划分为二手房成交信息数据和二手房挂牌信息数据；按照城市类型进行划分，展现了一线、新一线、二线、三四线城市的二手房成交与挂牌情况；
- **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个省级行政区、直辖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、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.全国新房数据

- 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；
- 含供地结果基本信息表、分期支付约定信息表、土地出租/抵押/转让表；
- **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、原及现土地使用权人等。

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，最终形成中国各省份官网报纸数据，以供学者对媒体问题进行细致研究；
- 字段：日期、版面、前标题、标题、副标题、图片、内容等。

2.中国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为主体字段；
- 包含帖子详情表、回帖详情表、用户详情表、用户粉丝表、用户关注表；
- 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

- 时间区间为2011-2021年，全面获取地方政府留言板问答的文本内容，数据量级过四百万，且未来仍会持续更新；
- 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【中国专利与创新数据】板块

5.CCTV新闻联播文本数据

- 时间区间为2016.06.15至2021.12.31，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.中国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。

7.中国上市公司全部公告数据

- 涵盖年报、半年报、三季报、公司治理、股东大会、日常经营等26个类型公告，信息量极全，对中国上市公司研究、证券市场研究等方面起到有力的数据支撑；
- 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，时间区间为2006.04至2021.12.31；
- 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；
- 包含诗词表和作者信息表等六张表格，多达25万行；
- 字段：唐诗、宋词、宋诗作者/名/正文等。

1. 中国上市公司财务信息数据

- 覆盖中国上市公司财务报表的所有指标，包含一季报、三季报、中报、年报等全部财务信息，以及业绩预告、所有者权益变动、财报披露时间等信息；
- **字段**：证券代码、公司名称、会计年度、营业收入、研发费用、营业利润、货币资金、应收账款、支付的各项税费、吸收投资收到的现金、业绩预告及内容、本期增加（股本）、本期减少（盈余公积）、期末数（未分配利润）、财报首次预约时间、一次变更日期、实际披露时间等。

2. 中国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，文本与数值型数据并存；
- **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. 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，最早可追溯至2013年；
- **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. 中国上市公司及子公司名称数据

- 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，并发布了历年上市公司及子公司名称数据；
- **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。

5. 中国上市公司及子公司基础信息数据

- 将上市公司及子公司名称数据与中国工商注册基本信息数据匹配，补足了上市公司及子公司在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、工商注册代码、地址、经营范围、经营状态等。

6. 中国上市公司高管信息数据

- 包含中国上市公司的全部在任高管信息及高管股权增减持信息；
- **字段**：证券代码、公司名称、高管姓名、职务、持股数、年薪、变动日期、变动股数、均价。

7. 中国上市公司内部交易数据

- 涵盖了中国上市公司高管人员之间股权交易的信息，与高管股权增减持数据不同，本数据更加侧重于上市公司的内部交易；
- **字段**：证券代码、公司名称、高管姓名、变动人、变动日期、变动量、成交均价、变动后持股数、与董监高关系、高管职务。

8.中国上市公司分红送配信息数据

- 涵盖了在中国上市公司各个报告期的分红送配信息；
- 字段：证券代码、上市公司、报告期、方案、股权登记日、除权除息日、派息日。

9.中国上市公司增发信息数据

- 涵盖了在中国上市公司各个报告期的增发股票信息；
- 字段：证券代码、上市公司、增发时间、价格、股数、募集资金。

10.中国上市公司股权信息数据

- 涵盖了在中国上市公司各个报告期的股本结构信息；
- 字段：证券代码、公司名称、股权变动日期、总股本、流通股本、非流通股本、限售股本等。

11.中国上市公司股东信息数据

- 涵盖了在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股东的持股信息、基金持股信息、限售解禁信息等；
- 字段：证券代码、公司名称、报告期、股东名称、持股数量、持股比例、较上期变动、机构名称、股东总人数、A/B/H股股东数、可流通时间、可流通数量、流通股份类型等。

12.中国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、现场投票股东数量、网络投票股东数量、赞成股份比例、反对股份比例等。

13.中国上市公司重要事项数据

- 涵盖了在中国上市公司在经营期间发布的重要事项；
- 字段：证券代码、会计年度、公司名称、公告日期、备忘事项、事件类型。

14.中国上市公司全部公告数据

- 见前【文本数据】板块

15.中国上市公司专利及引用被引用数据

- 见前【中国专利与创新数据】板块

16.中国上市公司绿色专利及引用被引用数据

- 见前【中国专利与创新数据】板块

17.中国上市公司股吧评论数据

- 见前【文本数据】板块

18.中国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。

19.中国财经报纸文本数据

- 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。

20.中国上市公司投资者关系管理数据

- 见前【文本数据】板块

21.中国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。

22.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，包含2020.01.01之前的所有数据，为案例分析、大样本实证研究提供了数据基础；
- 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。

23.中国上市公司招聘数据

- 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。

24.中国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。

25.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；依据证券机构类型，将人员基本信息表和人员职业经历变化表分别划分为证券公司、证券资产管理公司、证券投资咨询机构、证券资信评级机构四个部分；
- 字段：姓名、性别、执业机构、登记编号、学历、登记状态等。

26.基金从业人员信息数据

- 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、从业资格类别、证书取得日期、变更日期、证书状态等。

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。
- CnOpenData数据团队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，CnOpenData建立了本数据的展示区及数据索引，便于学者浏览。

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。
- CnOpenData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。
- CnOpenData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。

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。
- CnOpenData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-OPEN

- 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。
- CnOpenData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)-OPEN

- 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，CnOpenData建立了该数据的展示区及数据索引，便于学者浏览。

1.全球专利及引用被引用数据

- 囊括全球多个国家（如美国、日本、欧洲等国家或地区）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（授权）公布号、IPC分类号、引用及被引用专利名称等。

2.美国专利及引用被引用数据

- 囊括美国的发明、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。

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.英国专利及引用被引用数据

- 囊括英国的发明、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。

4.日本专利及引用被引用数据

- 囊括日本的发明、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.韩国专利及引用被引用数据

- 囊括韩国的发明、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。

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，明细表则对会议中产生的会议记录、决议草案、决议、委员会报告等文件做了详细分类与解读，并提供中英文双版文件下载地址；
- 字段：主题、标题、议程、决议、决定、投票、投票日期等。

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2015-2021年的政府采购合同信息；
- 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. 中国地方政府工作报告文本数据

- 覆盖近20年内全国29个省份及对应的304个地级市，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。

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. 失信企业基本信息数据增强版

- 见前【中国工商注册数据】板块

10. 中国行政区划数据-OPEN

- 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1987.01-2020.01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。

1. 中国各省份官方报纸数据

- 见前【文本数据】板块

2. 中国工商注册企业数量统计数据

- 见前【中国工商注册数据】板块

3. 中国各区县最低工资数据

- 各地区最低工资数据数据库，包含各区县最低工资划分标准（1999.06.01-2019.01.01）、金额（1999.07.01-2019.01.01）和调整日期；
- 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-OPEN

- 分为两张子表：历年县以上的行政区划信息（1981-2020）、历年县以下的行政区划变更信息（2014-2020）；
- 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. 中国各地区碳交易数据

- 见后【绿色数据】板块

1. 中国工业企业（绿色）专利及引用被引用数据

- 见前【中国专利与创新数据】板块

2. 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。

3. 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。

4. 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。

5. 中国工业企业基本信息数据

- 见前【中国工商注册数据】板块

1. 中国绿色专利申请与授权数据

- 见前【中国专利与创新数据】板块

2. 中国上市公司绿色专利及引用被引用数据

- 见前【中国专利与创新数据】板块

3. 中国工业企业绿色专利及引用被引用数据

- 见前【中国专利与创新数据】板块

4. 中国绿色专利分地区统计数据

- 见前【中国专利与创新数据】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-Firm-level Climate Change Exposure

- 见前【公共数据】板块

7. 30米全球地表覆盖数据-OPEN

- 见前【公共数据】板块

8. 空气质量站点监测数据-OPEN

- 见后【环境数据】板块

9. 中国历史天气数据-OPEN

- 见后【环境数据】板块

1. 中国上市公司高管信息数据

- 见前【中国上市公司数据】板块

2. 证券从业人员信息数据

- 见前【中国上市公司数据】板块

3. 基金从业人员信息数据

- 见前【中国上市公司数据】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、事务所编号、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。

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2014-2018年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，为区域经济、农村金融、银行经营方面的研究提供数据支持；
- 字段：省市区、服务网点名称、邮编、地址、营休时间、营业状态等。

1. 中国海关统计数据

- 整理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，包括18个子表，为研究中国进出口贸易相关问题提供权威、准确、全面的数据支持；
- 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

- 整理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，包括24个子表，以月度数据为基础，涵盖依据《商品名称及编码协调制度》（缩写HS，分为HS4位编码指数和HS2位编码指数）、《国际贸易标准分类》（缩写SITC，按初级产品、工业制品两大项及下一级子项分别计算其指数）、《按大类经济类别分类》（缩写BEC，分为资本品、中间产品和消费品）、国民经济行业分类（分为农林牧渔业、采矿业、制造业等）四种分类标准下，我国进口、出口商品贸易的月度同比、环比指数；
- 字段：HS编码、SITC编码、日期、参照基准、商品名称、价格指数等。

1. 台风数据

- 含台风基本信息、台风路径信息、台风风圈半径表；
- 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。

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；
- 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、周边城市等。

1. 人口迁徙大数据

- 包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。

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，该数据库收录了全国各公司捐款金额、捐款去向、捐款时间等信息；
- 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。

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。

业 医
数 疗
据 行

1. 中国在册船舶信息数据

- 根据CSA和CSAD分类，识别了国际与国内1975年至今在中国注册的船舶信息；
- 字段：船名、船舶归属类别、船东、船旗国、船籍港、船区等。

2. 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、工厂名称、公司名称、联系人地址等。

业 船
数 舶
据 行

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、经纬度等。

2. 中国各地区方言数据

数 文
据 化

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年后每年）、所有高铁线路表及各高铁站点开通时间表；
- 字段：列车编号、起始站编号名称、出发时间、车站编码、服务类型、途径站名称、高铁站开工及通车时间、高铁线路。

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。

交 通
行 业
数 据

1.全球摩天大楼数据

- 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、邮政编码等。

业建
数筑
据行

1.全国旅游景点信息数据

- 覆盖中国各地区旅游景点的详细信息；
- 字段：景点名称、景点地址、开放时间、门票价格、景点介绍、优待政策、服务设施、景区简介等。

2.全国酒店住宿信息数据

- 覆盖中国各地区各类型的酒店住宿的详细信息；
- 字段：酒店名称、位置、开业时间、装修时间、酒店简介、客房数、联系方式等。

旅
游
行
业
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

- 全面涵盖了电影相关的信息，尤其包括大量简洁的短篇以及优质的长篇电影评论；
- 字段：电影名称、豆瓣评分、评分ID、评分、评论内容、演员名称、出生日期及出生地等。

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

- 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。

5.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，以供学者对民族民俗文化等相关问题进行细致研究；
- 字段：上联、下联。

6.美国金球奖获奖数据

- 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、是否获奖等。

吃
喝
玩
乐
数
据

CnOpenData 的朋友

学术合作单位



业界合作机构



中国高校客户



海外高校客户



数据的顶端

也是学术的顶端

服务内容

线上平台已发布的数据

与客服联系获取公司部分私有数据

部分私有数据已经在国内外顶级期刊上发表多篇学术论文。

个性化数据定制服务

这是我们的老本行，您只需说明需求，我们评估后便可进行获取。
数据量越大、处理难度越高的数据，我们越有信心！



Q CnOpenData

公众号



Q CnOpenData-ybl

销售经理-杨宝璐



Q CnOpenData-wsw

销售经理-王诗薇



Q CnOpenData-dxy

机构服务-丁筱雨

官网链接：www.cnopendata.com

邮箱地址：wangshiwei@cnopendata.com

yangbaolu@cnopendata.com

dingxiaoyu@cnopendata.com

联系电话：+86 17720481357 杨宝璐

+86 13370290605 丁筱雨