



有机数中国环保排污数据

使用文档

2023年2月

•

目录

数据简介	3
数据内容	3
数据特色	5
数据采集	6
数据维护	6

数据简介

本档为有机数客户提供环保排污数据包的相关使用说明。有机数排污数据是有机数 ESG 数据系列的其中一个环境类基础数据包，为客户提供自 2014 年起的企业污染物排放数据。该数据包收集了全国各省市环保行政部门发布的排污许可证及环保处罚信息，覆盖排污企业超过 44 万家，其中包含上市公司近 1500 家及其下属子公司超过十万家，覆盖污染物超过 600 个种类。

覆盖全面：包含中国各省市地方环保局披露环保排污相关所有信息及历史数据

更新及时：数据获取流程自动执行，月度更新最及时排污信息

数据准确：大数据专家团队定制的数据处理模块，系统稳定，数据精度高

结构清晰：后期经过多轮行业专家分类处理，结构清晰，可用性高

数据内容

有机数环保排污数据集包含细节表及统计表两个数据表，数据库包含以下内容：

排污企业信息：

企业名称、社会信用代码、省份、生产所在地、行业、主要产品等属性信息；

排污企业对应的原始数据：

- 排污许可证：编号、行政区划、行业、发证日期、生效/失效时间等；
- 排污信息：污染物类别、种类、排放规律、执行标准等；
- 污染物监测数据：监测的时间、地点、项目、结果、浓度、阈值以及是否超标等；
- 环保处罚信息：处罚机关、处罚的事由、时间、详情等

有机数以上述信息为基础，经过专家的分类，整理出了上市公司排污浓度统计表。统计表的制作步骤如下：

1. 将超过 600 中污染物以两种方式进行了分类，第一种分类方式包括气体、液体、固体、菌类和其他等五种物理性质类别；第二种分类方式包括金属、烟尘、无机盐、有机物、生物类和其他等六种生化性质类别。
2. 使用有机数企业信息数据库，将最终排污企业与上市公司进行匹配，得到上市公司旗下所有相关排污单位列表。
3. 以月为单位，统计上市公司在该月旗下所有排污数据的浓度均值、中位数及环比变化率。

综上，统计表包含了月度上市公司旗下子公司污染物排放的分类统计数据，统计方法可选平均值、中值以及环比变化率。

环保排污细节表（仅国内客户可用）：

id	serial	数据 ID
ticker	text	上市公司代码
parent_company	text	母公司名称
pollution_company	text	排污企业名称
monitor_date	timestampz	监测日期
monitor_place	text	监测地点
monitor_project	text	监测项目
new_monitor_result	float4	监测结果
new_unit	text	排放物单位
emission_standard	text	排放标准
emission_entry	text	排污标准文件
standard_limit	text	排污标准限值
is_exceed_standard	bool	是否超限
excessive_multiple	float4	超限倍数
new_category	text	污染物类别 1
project_class	text	污染物类别 2
release_time	timestampz	公布日期
backtest_date	timestamp	回测日期
result_pct_chg	float4	结果环比变化
created_at	timestampz	创建日期
updated_at	timestampz	更新日期

deleted_at	timestampz	删除日期
------------	------------	------

环保排污统计表:

字段名	字段属性	字段含义
id	serial	数据 ID
ticker	text	上市公司代码
backtest_date	timestamp	回测日期
all	float4	所有污染类别
other_category	float4	其他大类 1
gas	float4	气体
liquid	float4	液体
fungi	float4	菌类
solid	float4	固体
other_class	float4	其他大类 2
metal	float4	金属
smoke	float4	烟尘
inorganic_salt	float4	无机盐
organic	float4	有机物
biological	float4	生物类
backtest_category	text	计算方式（平均、中值、环比）
created_at	timestampz	创建日期
updated_at	timestampz	更新日期
deleted_at	timestampz	删除日期

数据特色

有机数环保排污数据集是国内领先的企业排污数据库，多项独特设计为机构投资者提供高质量另类数据：

- 人工智能赋能数据清洗，可用性高
- 清洗后数据保留最细颗粒度，不丢失任何细节

- 行业专家分类统计，方便进行因子回测
- 数据抓取清理全自动化

数据采集

排污数据来源为各省市环保局、环境监督机关、政府数据开放平台等官方监测信息，通过自然语言处理技术，将信息格式为网页文本、网页表格、word/PDF 文件、Excel 表格的原始文档转换为可辨认的数据格式，对于采集的各类信息进行解析、甄别、分类、入库，采用实体消歧、指代消解等知识图谱技术，使用人工智能模型识别公司主体，打通数据链。

原始数据清洗完成后，有机数的行业专家针对不同污染物的性质，将超过 600 中不同的污染物分成 9 大类别，以便于量化团队回测及行业研究员进行分析。

其后有机数量化分析团队会对完成后的数据集进行量化回测，将回测结果反馈至数据团队，并与数据团队共同优化数据处理。

数据维护

每月定期数据采集流程后有数据质量校验模块进行自动检验，在监测到异常时会呼叫人工介入，校验数据，提高数据精度。

有机数

北京市海淀区上地六街 26 号 410 室

Email: ESG@youjinvest.com

Web: www.youjinvest.com

版本号 1.3 发布日期 2022 年 4 月