

Web Scraping Policy



Web scraping policy

Scope

This policy is applicable to all Office for National Statistics (ONS) staff activities involving web scraping of personal and non-personal data. Web scraping is defined as the collection of data automatically retrieved from the internet. When obtaining or procuring web-scraping services from a third party, the ONS will seek to ensure that the overarching principles contained in this policy are met. The rest of this policy outlines the main principles of web scraping and provides practical guidance.

This policy does not cover the use of Application Programming Interfaces (APIs). This policy also differs from the ONS Social Media Data Policy, which outlines procedures related to the collection, use and analysis of social media data obtained from social media platforms. Neither is this policy the same as the ONS Open Data Policy, which provides guidance on the collection and use of open government, academic, for-profit and non-profit organisation data for statistics and statistical research.

Purpose

Use of alternative data sources is an important element of the ONS's current five-year strategy, [Statistics for the Public Good](#), for delivering high-quality data and analysis to inform the UK, improve lives and build the future. Driven by this strategic imperative, ONS staff may use web scraping as an alternative data collection mechanism that can complement and improve traditional forms of data collection, such as surveys.

The purpose of this policy is to ensure that web scraping at the ONS is carried out transparently, consistently, ethically, and in accordance with all relevant legislation.

Background

This policy sets out the practices and principles that ONS staff will follow when scraping data from websites to produce statistics and conduct statistical research, including exploratory research, which serves the public good.

When web scraping, we will ensure that we minimise any burden on websites, respect the Robots Exclusion Protocol and associated restriction, abide by all applicable legislation, and monitor the evolving legal situation.

Policy statement

Web scraping is only conducted by the ONS for the purposes of any one or more of its functions set out in the Statistics and Registration Service Act 2007 and Census Act 1920.

These Acts limit the functions of the ONS to the production and publication of official statistics that serve the public good.

We will adopt the following overarching principles to guide our approach to web scraping:

- minimise burden on website owners
- respect the Robots Exclusion Protocol
- abide by all applicable legislation and monitor the evolving legal situation

Roles and Responsibilities

ONS staff who request web scraping

Staff who request web scraping must comply with the ONS Web Scraping Policy and consult with the Data Acquisition team before commencing any web-scraping activities. They are responsible to their line managers.

Data Acquisition, Data Acquisition and Operations (DAO)

The DAO team must advise ONS staff on any alternative and/or existing data sources, ensuring that web scrapers are fully compliant with the Web Scraping Policy. They must also receive and evaluate the web scraping request from ONS staff, seek advice from ONS Legal Services, the UK Statistics Authority (UKSA) Data Ethics team, and/or the Data Governance Committee (DGC) when needed. They must also engage with the website owners' opt-out requests and any enquiries, while monitoring and keeping all records of ONS web scraping activities. They are responsible to Data Governance Committee.

ONS Legal Services

ONS legal services provide advice and guidance on current and evolving legal context of open data

UKSA Data Ethics

The UKSA Data Ethics team provides advice on ethical issues as the first point of contact, as well as the National Statistician's Data Ethics Advisory Committee (NSDEC) if additional advice from NSDEC is required. They are responsible to the Head of Data Governance Policy and Legislation.

Data Governance Committee (DGC)

The DGC ensures the consistent application of this policy to all ONS staff, and advises and assesses the organisational risk of conducting web scraping. They are responsible to the National Statistics Executive Group.

National Statistician's Data Ethics Advisory Committee (NSDEC)

NSDEC provides independent advice on ethical issues if required. NSDEC is responsible to the National Statistician.

Supporting documents

Annex 1. Guidance for website owners

This document includes available options for website owners and how to opt out