

# CESSDA Data Catalogue Overview

June 2018

John Shepherdson  
Chair, CESSDA Technical Working Group

# CESSDA Data Catalogue (CDC)

Will launch in June 2018

Does not hold any data itself, or provide direct access to data

Automatically, frequently and periodically harvests metadata from Service Providers and other sources

Fully automated harvesting is crucial to ensure that the CESSDA Data Catalogue content is up to date

Researchers are directed to the source/service where the data may be obtained or analysed

See also <https://www.cessda.eu/Research-Infrastructure/Products-Services>

cessda eric

# CESSDA Data Catalogue

Circa 13,000 records to date:

FSD

GESIS

Progedo

UKDS

Actively working with other Service Providers

cessda eric

# Basic Features

Consortium of European Social Science Data Archives

Language: English ▾



Find Social and Economic Research Data

13260 results found in 124ms

Advanced search

Reset filters

Clear search

► Topic



► Collection years



▼ Country



Search countries

▼ Publisher



Search publishers

▼ Language of data files



Search languages

Results per page

30 ▾

Sort by

Title (ascending) ▾



1

2

3

4

...



[Perspective Germany 2005/06](#)

*McKinsey & Company, Düsseldorf; Stern, Hamburg; Zweites Deutsches Fernsehen (ZDF), Mainz; (1 more)*

[Applicant Countries Eurobarometer 00 \(pilot study\)](#)

*European Commission, Brussels DG Education and Culture (EAC-D2, Public Opinion Analysis)*

Attitude towards the EU and EU enlargement. Topics: life satisfaction; subjective rating of the development of the general life situation, the economic situation, the financial situation of the household, personal job situation, national labor market situation and the personal professional outlook in the coming year; native language; knowledge of foreign languages; frequency of political discussions with friends; self-rated opinion leadership; frequency of news consumption (television, newsp...

[Flash Eurobarometer 163 \(0 employee = One-person-enterprises\)](#)


*Papacostas, Antonis (European Commission, Brussels DG Communication Public Opinion Analysis Sector)*

cessda eric

# Help - Tooltips

Consortium of European Social Science Data Archives

Language: English ▾

 **cessda**  
**DC Data Catalogue**

13260 results found in 124ms

Advanced searchReset filtersClear search

► Topic?

► Collection years

▼ Country

▼ Publisher?

▼ Language of data files?

Search countries

Search publishers

Search languages

CESSDA Topic Classification serves to identify the general topics, subjects or themes of a study.

Results per page30

Sort byTitle (ascending)

<

1

2

3

4

...

>

[Perspective Germany 2005/06](#)

McKinsey & Company, Düsseldorf; Stern, Hamburg; Zweites Deutsches Fernsehen (ZDF), Mainz; (1 more)

[Applicant Countries Eurobarometer 00 \(pilot study\)](#)

European Commission, Brussels DG Education and Culture (EAC-D2, Public Opinion Analysis)

Attitude towards the EU and EU enlargement. Topics: life satisfaction; subjective rating of the development of the general life situation, the economic situation, the financial situation of the household, personal job situation, national labor market situation and the personal professional outlook in the coming year; native language; knowledge of foreign languages; frequency of political discussions with friends; self-rated opinion leadership; frequency of news consumption (television, newsp...

[Flash Eurobarometer 163 \(0 employee = One-person-enterprises\)](#)

Papacostas, Antonis (European Commission, Brussels DG Communication Public Opinion Analysis Sector)

▼ Topic



Search topics

▼ Collection years



1900



2018

1900

-

2018

Go

▼ Country



Search countries

▼ Publisher



Search publishers

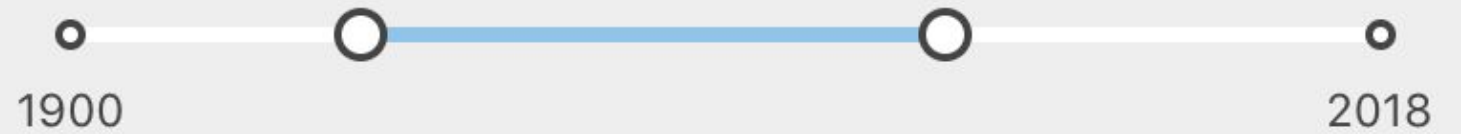
▼ Language of data files



Search languages

cessda eric

▼ Collection years



cessda eric

▼ Topic?

▼ Collection years?

1900

2018

1900

-

2018

Go

▼ Country?

▼ Publisher?

▼ Language of data files?

Filter summary

Advanced search

Reset filters

Clear search



## Filter summary



The following filters have been applied to your search.

**Topic:** medical behavior

**Collection years:** 1942 - 2018

**Country:** England Finland

**Language of data files:** fi

Select a filter to remove it from this search.

Close

# Uses Deep Links

[https://datacatalogue-dev.cessda.eu/?publisher.publisher\[0\]=UK%20Data%20Service&publisher.publisher\[1\]=GESIS%20Data%20Archive%20for%20the%20Social%20Sciences&publisher.publisher\[2\]=Finnish%20Social%20Science%20Data%20Archive&publisher.publisher\[3\]=Centre%20de%20Donn%C3%A9es%20Socio-Politiques%20%28CDSP%29](https://datacatalogue-dev.cessda.eu/?publisher.publisher[0]=UK%20Data%20Service&publisher.publisher[1]=GESIS%20Data%20Archive%20for%20the%20Social%20Sciences&publisher.publisher[2]=Finnish%20Social%20Science%20Data%20Archive&publisher.publisher[3]=Centre%20de%20Donn%C3%A9es%20Socio-Politiques%20%28CDSP%29)

# Results - Pagination

Consortium of European Social Science Data Archives

Language: English ▼



Find Social and Economic Research Data

13260 results found in 124ms

Advanced search

Reset filters

Clear search

▼ Topic



Search topics

Results per page

10  
✓ 30  
50  
150

Sort by

Title (ascending) ▼



1

2

3

4

...



▼ Collection years



[Perspective Germany 2005/06](#)

McKinsey & Company, Düsseldorf; Stern, Hamburg; Zweites Deutsches Fernsehen (ZDF), Mainz; (1 more)



cessda eric

# Results - Sort Order

Consortium of European Social Science Data Archives

Language: English ▾



Find Social and Economic Research Data

13260 results found in 124ms

Advanced search

Reset filters

Clear search

▼ Topic



Search topics

▼ Collection years



Results per page

30 ▾

Sort by

Relevance

✓ Title (ascending)

Title (descending)

Date of collection (ascending)

Date of collection (descending)



1

2

3

4

...



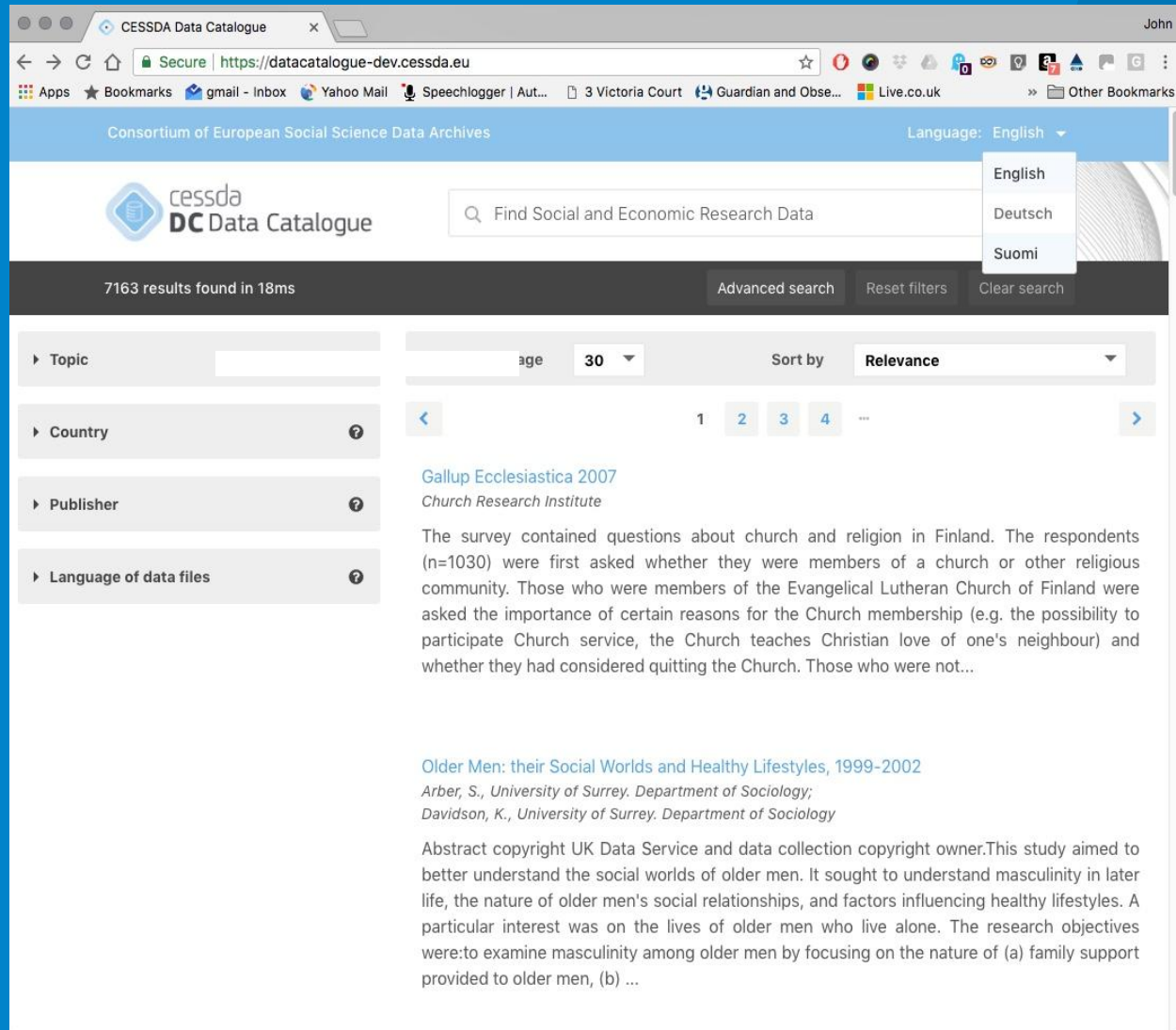
[Perspective Germany 2005/06](#)

McKinsey & Company, Düsseldorf; Stern, Hamburg; Zweites Deutsches Fernsehen (ZDF), Mainz; (1 more)



cessda eric

# Select UI Language



The screenshot shows the CESSDA Data Catalogue website. The browser address bar displays 'https://datacatalogue-dev.cessda.eu'. The page header includes the CESSDA logo and the text 'Consortium of European Social Science Data Archives'. A search bar contains the text 'Find Social and Economic Research Data'. Below the search bar, it indicates '7163 results found in 18ms'. A language selection dropdown menu is open, showing options for 'English', 'Deutsch', and 'Suomi'. The main content area displays search filters on the left (Topic, Country, Publisher, Language of data files) and search results on the right. The first result is 'Gallup Ecclesiastica 2007' by the Church Research Institute, with a description of the survey. The second result is 'Older Men: their Social Worlds and Healthy Lifestyles, 1999-2002' by Arber, S., and Davidson, K., with a description of the study.

Language: English ▼

- English
- Deutsch
- Suomi

7163 results found in 18ms

Advanced search Reset filters Clear search

Topic

Country

Publisher

Language of data files

age 30 Sort by Relevance

1 2 3 4 ...

[Gallup Ecclesiastica 2007](#)  
Church Research Institute

The survey contained questions about church and religion in Finland. The respondents (n=1030) were first asked whether they were members of a church or other religious community. Those who were members of the Evangelical Lutheran Church of Finland were asked the importance of certain reasons for the Church membership (e.g. the possibility to participate Church service, the Church teaches Christian love of one's neighbour) and whether they had considered quitting the Church. Those who were not...


[Older Men: their Social Worlds and Healthy Lifestyles, 1999-2002](#)  
Arber, S., University of Surrey. Department of Sociology;  
Davidson, K., University of Surrey. Department of Sociology

Abstract copyright UK Data Service and data collection copyright owner. This study aimed to better understand the social worlds of older men. It sought to understand masculinity in later life, the nature of older men's social relationships, and factors influencing healthy lifestyles. A particular interest was on the lives of older men who live alone. The research objectives were: to examine masculinity among older men by focusing on the nature of (a) family support provided to older men, (b) ...

# Select UI Language

Consortium of European Social Science Data Archives

Kieli: Suomi ▼

 cessda  
DC Data Catalogue

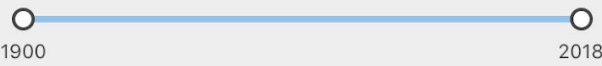
Hae yhteiskunta- ja taloustieteiden tutkimusaineistoja

1373 tulosta ajassa 36ms

Tarkennettu haku Poista suodattimet Tyhjennä haku

**Aihepiiri** ?

**Aineistonkeruun ajankohta** ?

  
1900 - 2018 Go

**Aineiston keruumaa** ?

Hae aineistojen keruumaita

Tuloksia per sivu 30 ▼

Lajittele Otsikko (nouseva) ▼

< 1 2 3 4 ... >

[Sosiaalilääketieteilijän identiteetti 1999](#)  
*Aukee, Ranja (Tampereen yliopisto. Sosiaalipolitiikan ja sosiaalityön laitos)*

Sosiaalilääketieteen yhdistyksen jäsenille 1990-luvun lopulla suunnatun kyselyn avulla oli tarkoitus selvittää, millainen käsitys heillä on Sosiaalilääketieteellisestä yhdistyksestä, sosiaalilääketieteestä ja sosiaalilääketieteilijöiden identiteetistä. Lisäksi kyselyssä oltiin kiinnostuneita, keitä sosiaalilääketieteilijät pitivät alansa merkittävinä ammattilaisina. Aineistossa käytössä on huomiotava, että avokysymysten vastaukset eivät sisälly aineistoon muutamaa poikkeusta lukuun ottamatta...

[Advanced search](#)[Reset filters](#)[Clear search](#)**▼ Similar results**[Television and Press Coverage of the 1997 British General Election](#)[British General Elections, 1955-1979](#)[British General Election Study, 1997 : Campaign Panel](#)[Content Analysis of Newspaper Coverage, British General Election, 2001](#)[Television: The Public's View, 1997](#)[< Back](#)[</> View JSON](#)[🔗 Go to study](#)**Study title**

Effects of Television News in British General Elections, 1997

**Creator**

Sanders, D., University of Essex. Department of Government

Norris, P., Harvard University. John F. Kennedy School of Government

**Study Persistent Identifier**

4167 (UKDA)

10.5255/UKDA-SN-4167-1 (datacite)

**Abstract**

Abstract copyright UK Data Service and data collection copyright owner.

The research used an experimental design to assess the extent to which UK voters' political perceptions are shaped by the content of television news coverage. The research involved:

1. administering a pre-test questionnaire to a randomly selected sample of 1125 respondents;
2. randomly assigning respondents to a series of groups;
3. exposing each group of respondents to a distinctive selection of video news items; and
4. administering a post-test questionnaire to each respondent.

The purpose of the research was to establish the extent to which changes between pre-test and post-test responses varied according to the type of video footage that respondents had seen.

Three core sets of hypotheses were tested:

# Similar Results

## ▼ Similar results

[Gallup Ecclesiastica 1995](#)

[Gallup Ecclesiastica 2003](#)

[Gallup Ecclesiastica 1999](#)

[Gallup Ecclesiastica 2011](#)

[Gallup Polls, January 1974](#)

[< Back](#)

### Study title

Gallup Ecclesiastica 2007

### Creator

Church Research Institute

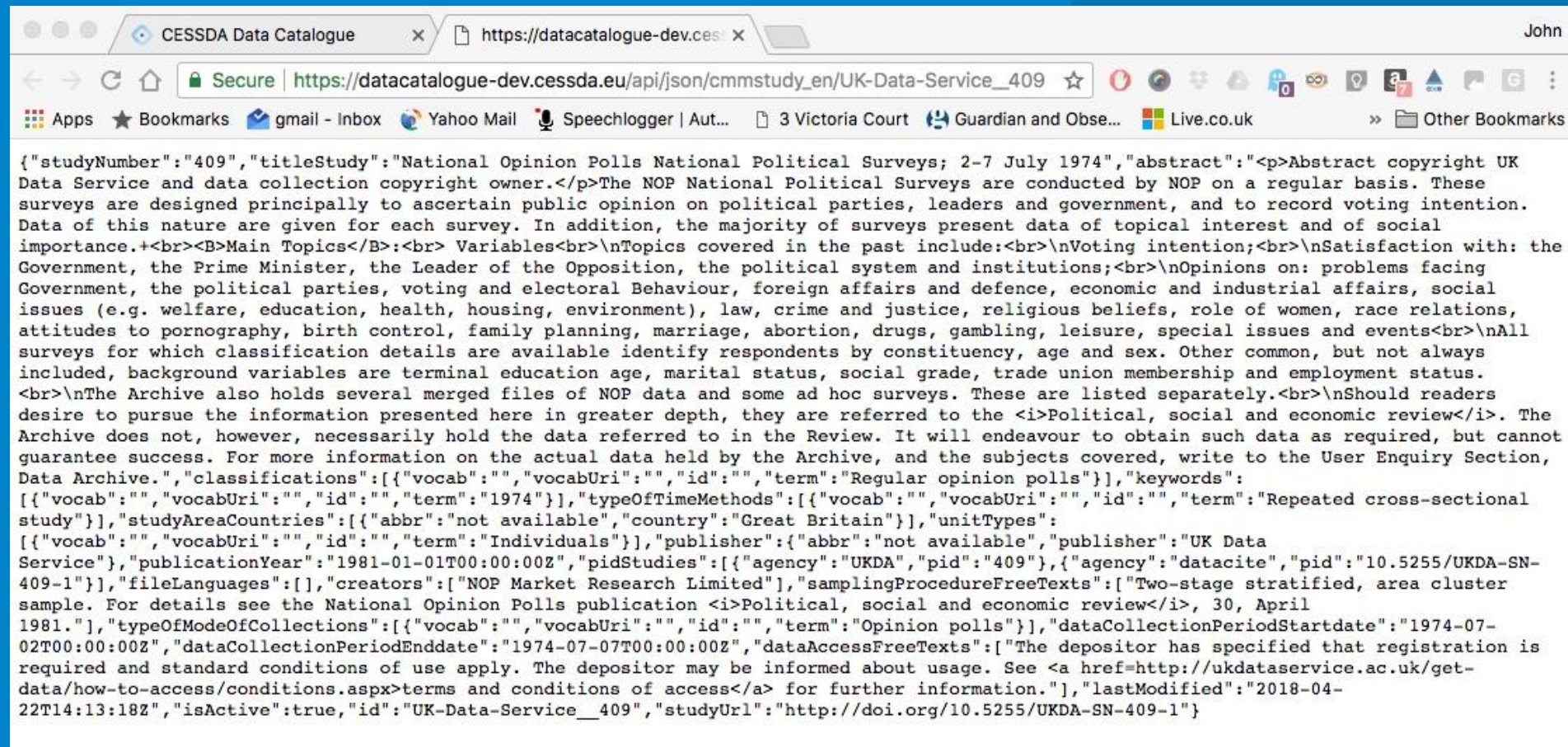
### Study Persistent Identifier

FSD2710 (FSD)

urn:nbn:fi:fsd:T-FSD2710 (The National Library of Finland)



# View JSON representation



The screenshot shows a web browser window with the address bar displaying `https://datacatalogue-dev.CESSDA.eu/api/json/cmmstudy_en/UK-Data-Service_409`. The page content is a JSON object representing a study record. The JSON includes fields for study number, title, abstract, classifications, keywords, study area countries, publication year, and data collection period. The abstract text is visible within the JSON structure, describing the National Opinion Polls National Political Surveys conducted by NOP on a regular basis. The JSON is formatted with line breaks and indentation for readability.

```
{
  "studyNumber": "409",
  "titleStudy": "National Opinion Polls National Political Surveys; 2-7 July 1974",
  "abstract": "Abstract copyright UK Data Service and data collection copyright owner. The NOP National Political Surveys are conducted by NOP on a regular basis. These surveys are designed principally to ascertain public opinion on political parties, leaders and government, and to record voting intention. Data of this nature are given for each survey. In addition, the majority of surveys present data of topical interest and of social importance. Main Topics: Variables Topics covered in the past include: Voting intention; Satisfaction with: the Government, the Prime Minister, the Leader of the Opposition, the political system and institutions; Opinions on: problems facing Government, the political parties, voting and electoral Behaviour, foreign affairs and defence, economic and industrial affairs, social issues (e.g. welfare, education, health, housing, environment), law, crime and justice, religious beliefs, role of women, race relations, attitudes to pornography, birth control, family planning, marriage, abortion, drugs, gambling, leisure, special issues and events. All surveys for which classification details are available identify respondents by constituency, age and sex. Other common, but not always included, background variables are terminal education age, marital status, social grade, trade union membership and employment status. The Archive also holds several merged files of NOP data and some ad hoc surveys. These are listed separately. Should readers desire to pursue the information presented here in greater depth, they are referred to the Political, social and economic review. The Archive does not, however, necessarily hold the data referred to in the Review. It will endeavour to obtain such data as required, but cannot guarantee success. For more information on the actual data held by the Archive, and the subjects covered, write to the User Enquiry Section, Data Archive.",
  "classifications": [
    {
      "vocab": "",
      "vocabUri": "",
      "id": "",
      "term": "Regular opinion polls"
    }
  ],
  "keywords": [
    {
      "vocab": "",
      "vocabUri": "",
      "id": "",
      "term": "1974"
    },
    {
      "vocab": "",
      "vocabUri": "",
      "id": "",
      "term": "Repeated cross-sectional study"
    }
  ],
  "studyAreaCountries": [
    {
      "abbr": "not available",
      "country": "Great Britain"
    }
  ],
  "unitTypes": [
    {
      "vocab": "",
      "vocabUri": "",
      "id": "",
      "term": "Individuals"
    }
  ],
  "publisher": {
    "abbr": "not available",
    "publisher": "UK Data Service"
  },
  "publicationYear": "1981-01-01T00:00:00Z",
  "pidStudies": [
    {
      "agency": "UKDA",
      "pid": "409"
    },
    {
      "agency": "datacite",
      "pid": "10.5255/UKDA-SN-409-1"
    }
  ],
  "fileLanguages": [],
  "creators": ["NOP Market Research Limited"],
  "samplingProcedureFreeTexts": ["Two-stage stratified, area cluster sample. For details see the National Opinion Polls publication Political, social and economic review, 30, April 1981."],
  "typeOfModeOfCollections": [
    {
      "vocab": "",
      "vocabUri": "",
      "id": "",
      "term": "Opinion polls"
    }
  ],
  "dataCollectionPeriodStartdate": "1974-07-02T00:00:00Z",
  "dataCollectionPeriodEnddate": "1974-07-07T00:00:00Z",
  "dataAccessFreeTexts": ["The depositor has specified that registration is required and standard conditions of use apply. The depositor may be informed about usage. See a href=http://ukdataservice.ac.uk/get-data/how-to-access/conditions.aspx>terms and conditions of access</a> for further information."],
  "lastModified": "2018-04-22T14:13:18Z",
  "isActive": true,
  "id": "UK-Data-Service_409",
  "studyUrl": "http://doi.org/10.5255/UKDA-SN-409-1"
}
```

# Metadata Issues

Many and various, including:

- Leading and trailing spaces or non-printing characters
- Capitalisation
- Missing/incorrect language tags
- Date and time formats
- Alternative accepted spellings (e.g. Surinam, Suriname\*)

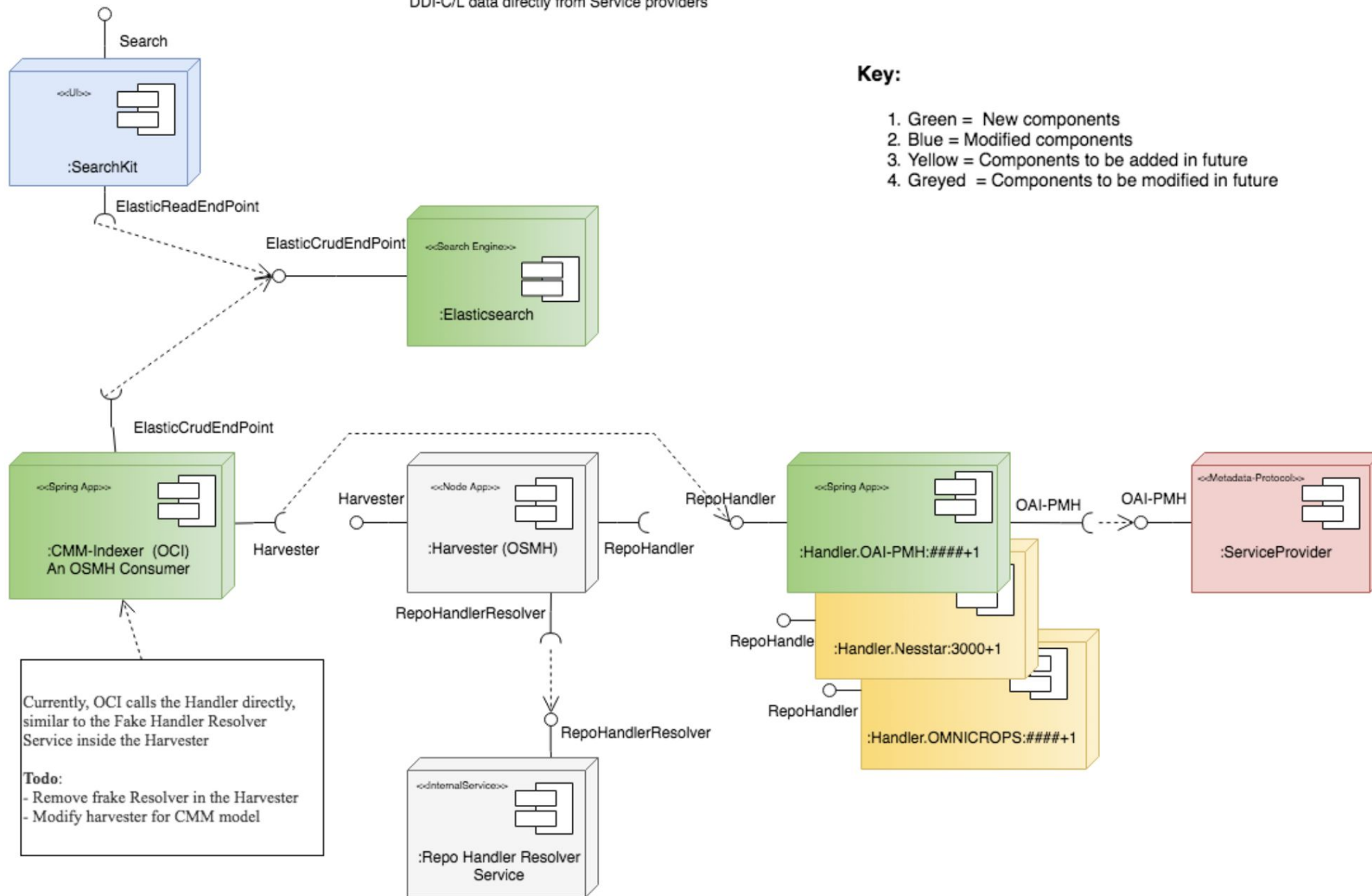
\*State in South America, formally referred to as Republic of Suriname

# Technical Details

- Java-based
- Spring Boot framework
- Elastic search
- UI - Javascript & Searchkit
- Docker containers running in Kubernetes clusters on Google Cloud Platform

# Technical Details

- Uses CESSDA's Open Source Metadata Harvester
- Supports CESSDA Metadata Model v1
- Repository Handlers
  - adapter for OAI-PMH (DDI XML 2.5)
  - adapter for NESSTAR (DDI XML 1.2)



# Version 1 (December 2017)

Fixed and upgraded Open Source Metadata Harvester

- Crawls endpoints for XML metadata and ingests transformed data into Elasticsearch

Leveraging a Microservice architecture with Spring Boot, Java 8 Stream for a declarative sets of business instructions:

- Harvester Microservice: Sole responsibility is to harvest and store XML metadata



# Version 1(December 2017)

Indexer Microservice: An ETL tool responsible for extraction of harvested metadata, transformation with named-entity recognition (NER) enrichment and ingestion into Elasticsearch.

Spring Boot Admin App: To monitor and manage the above Microservice's logs, configuration, JMX exported beans, fail fast feature toggles, etc.

Search UI: Worked with a Frontend developer to customise SearchKit for FullText Search with advanced search query and result filtering.

# Version 2 (June 2018)

Similar functionality to Version 1 - total rebuild to handle currently and future XML Schemas.

- Adopted a pluggable Microservice architecture style with multiple Repository Handlers, exposed as Restful APIs.
- Harvester responsible for discovering handlers and handing over requests to a given handler to process its XML schema specifics.
- Consumer Service responsible for orchestrating harvesting of any XML Metadata endpoint and ingestion of output into Elasticsearch.



# Version 3 (Date TBC)

- More metadata cleaning, transformation, disambiguation
  - Support for CESSDA Metadata Model v2
  - Harvest from other sources
  - Repository Handler adapter for OAI-PMH (DDI XML 3.2)
  - Make metadata harvestable by other aggregators
- 
- Identify relationships among all research entities  
publications, data, funding, researchers, organisations and data sources

# Thanks for listening

## Any Questions?

cessda eric