



Talking with data

Stories and lessons from my data adventures

Giuseppe Sollazzo • @puntofisso • csv,conf,v7



<https://puntofisso.net/csvconf2023>



Talking with data

Stories and lessons from my data adventures

NHS AI Lab Skunkworks

Demonstrating the potential for AI in health and social care

[What we do](#)

Read about our AI investigation

[Working with you](#)

Have you got an idea for using AI in health and social care?

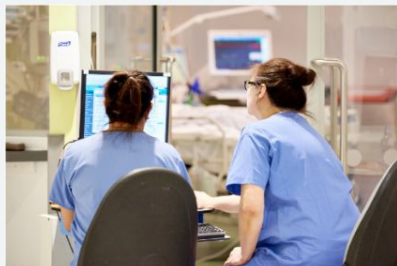
[Access the code](#)

We provide open source code from NHS AI Lab Skunkworks projects so that others can explore and develop AI tools.

National Health Service, 2020



The AI Skunkworks programme is part of the [NHS AI Lab](#). It finds new ways to use AI for driving forward the early adoption of technology to support health, in both clinical and business contexts. The team provides free short-term expertise and resources to public sector health and social care organisations to support AI projects and develop capability.



[Case study: Using AI to find optimal placement schedules for nursing](#)



[Case study: Creating synthetic data](#)

How we generated mock data



[Case study: Bed management](#)

An investigation into using AI to

Source
<https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/skunkworks>

Department for Transport, 2018

Department for Transport digital

Organisations: [Department for Transport](#)



Department
for Transport

blog



Introducing the Developing Data Unit – come and join us!

[Giuseppe Sollazzo](#), [Head of the Developing Data Unit](#), 11 February 2019 - [Data](#), [Digital](#), [Open Data](#), [Transformation](#)



About the DfT Digital blog

Read about things related to digital, technology and data happening in the Department for Transport (DfT) and with its partners.

[Find out more](#)

Follow us

[DfT Flickr](#)

Sign up and manage

Source
<https://dftdigital.blog.gov.uk/2019/02/11/introducing-the-developing-data-unit-come-and-join-us/>

St George's, University of London, 2009



Noemalife, 2007



Image CC BY-SA
https://www.flickr.com/photos/luca_volpi/2086483944/

Side gigs & data newsletter

QUANTUM OF SOLLAZZO

THE DATA NEWSLETTER BY @PUNTOFISSO

Quantum of Sollazzo is the weekly newsletter about data journalism, data visualization, tools, tutorials, data thinking, and AI. Basically, a weekly "best of" in all things data. Subscribe here and you will get it every Tuesday.

Enter your e-mail and join 3,400+ subscribers

ARCHIVES

- [Issues from 411](#)
- [Issues 113 to 410](#)
- [Issues 0 to 112](#)

Source
<http://puntofisso.net>



Images CC BY-SA
Giuseppe Sollazzo



Data advocacy

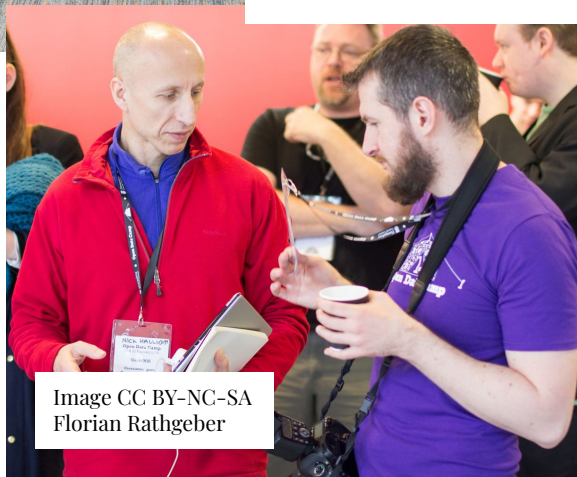


Image CC BY-NC-SA
Florian Rathgeber



Image CC BY-SA
ODI/Paul Clarke

Just imagine the effect that an army of
armchair auditors is going to
have on those expense claims.

David Cameron, 25 May 2009



Secretary of State's Code of Practice (datasets) on the discharge of public authorities' functions under Part 1 of the Freedom of Information Act

Issued under Section 45 of the
Freedom of Information Act

July 2013



The Freedom of Information (Release of Datasets for Re-use) (Fees) Regulations 2013

[◀ Previous](#) [Next: Provision ▶](#)

STATUTORY INSTRUMENTS

2013 No. 1977

FREEDOM OF INFORMATION

The Freedom of Information (Release of Datasets for Re-use) (Fees) Regulations 2013

Made

7th August 2013

Laid before Parliament

9th August 2013

Coming into force

1st September 2013

The Secretary of State makes the following Regulations in exercise of the powers conferred upon him by section 11B(1), (2) and (3) of the Freedom of Information Act 2000⁽¹⁾ and with the consent of the Treasury:

⁽¹⁾ [2000 c. 36](#). Section 11B was inserted by section 102(3) of the Protection of Freedoms Act 2012 ([c. 9](#)) ("the 2012 Act").

Form

Breakthrough Fund and Release of Data Fund

Information about the Release of Data Fund and Breakthrough Fund, and details of how to apply.

From: [Cabinet Office](#) and [Department for Business, Innovation & Skills](#)

Published 1 April 2014

Last updated 3 July 2014 — [See all updates](#)



Open Data User Group, 2012 – 2015

Sources

<http://odug.org.uk/members/>

<https://www.gov.uk/government/publications/breakthrough-fund-and-release-of-data-fund>



Data request title *

Title length limited to 60 characters, remaining: **60**

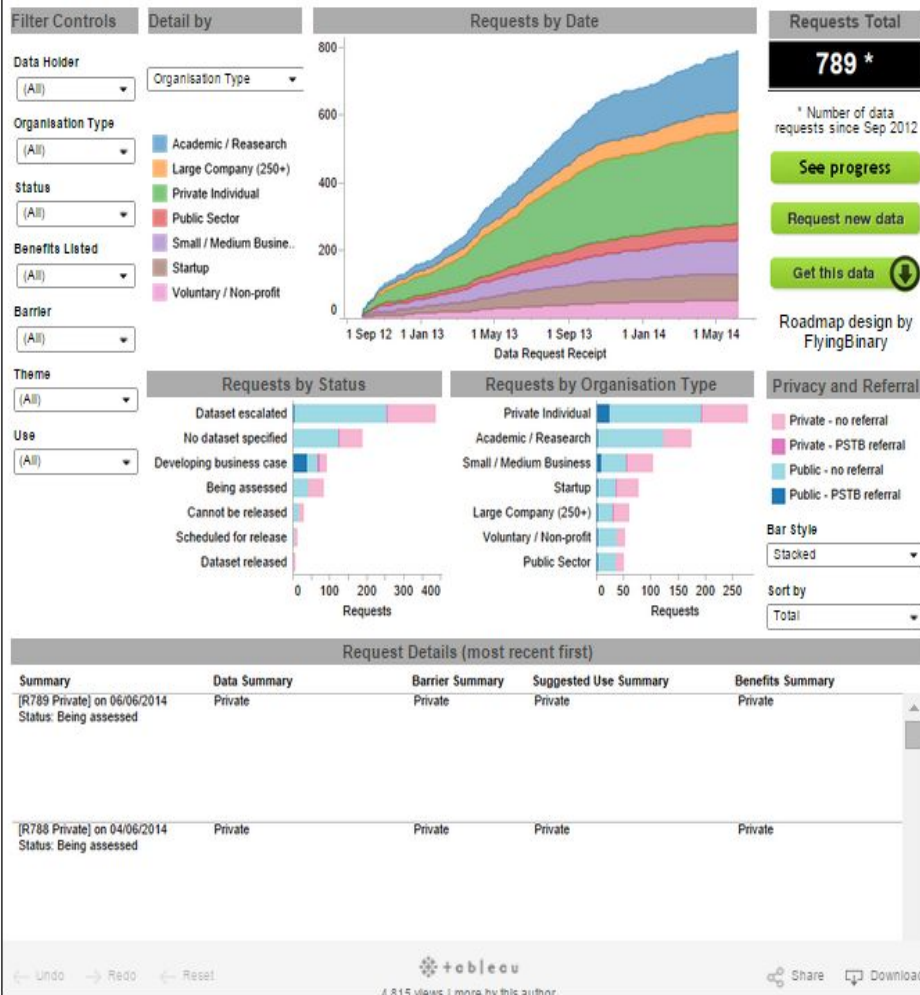
Data request description *

Publication preference *

- ☐ Request is public
- ☐ I would like my entire request to remain confidential

I request this data *

- ☐ As a private individual
- ☐ On behalf of an Organisation



Just imagine the effect that an army of
armchair auditors is going to
have on those expense claims.

David Cameron, 25 May 2009

The reality



CC BY-NC
<https://www.flickr.com/photos/nerdcoreblog/9444266896>



The Great British
Public Toilet Map

Vote People's Award 2016



We've been nominated!
Vote or tweet #T4Gtoiletmap

Add a toilet

Nearest Toilets



Source
<https://greatbritishpublictoiletmap.rca.ac.uk/>

Who owns and updates toilet data?

Local Authorities don't know whether to assign this task to their **(Open) Data people**, who don't know much about toilets, or to their **toilets people**, who don't understand much about data issues.

Gail Ramster, @gaillyk
Creator of the Great British Public Toilet Map

Lesson 1

Users have varied
data needs

Lesson 2

Open data works well
when it's part of a process
that collects and uses data
for a purpose

Authoritative dataset of GP and Dental Practices

July 2014

Name of dataset: GP and Dental Practices

1. Summary:

A complete and authoritative dataset of medical practices is currently missing from data.gov.uk. Having a list of medical practices is vital to foster excellence in health care on several dimensions, for example to:

- [give patients an informed freedom of choice
- [help health care professionals locate each other
- [allow businesses, software developers, data analysts and activists to build solutions to improve services

There are other sources where such data can be found, such as the Health and Social Care Information Centre (HSCIC) or the NHS Choices datasets pages; however, none of these sources is authoritative or has a clearly defined process for keeping the dataset up to date.

ODUG recommends that:

- [*the Department of Health establish an ongoing process to build, update and maintain on data.gov.uk an authoritative dataset of medical practices and operating practitioners, drawing on the datasets made available by HSCIC and NHS Choices*
- [*the dataset lists both GP and Dental practices and include data such as, but not limited to, practice details, opening times, location, e-mail address, status (for closed practices), telephone number*
- [*the dataset contains the patient acceptance criteria for each practice (for example, the catchment area where applicable)*
- [*the dataset contains a list of practitioners, so that a link between the practice and the practitioner(s) can be established*
- [*the dataset be refreshed monthly*
- [*the dataset provide a stable unique identifier for each practice, so that they can be referred to unambiguously in other datasets; Government guidelines recommend that this should be a URL (although not necessarily a URL that can be resolved)*
- [*the dataset should be ODI Open Data Certified and a Standard Level certificate be attained within 6 months*

Source

<https://odug.org.uk/authoritative-dataset-of-gp-and-dental-practices/>

An authoritative
dataset of doctors'
practice locations



Sam

@smithsam

Replying to @puntofisso

@puntofisso I'm unclear what @odugUK is actually doing on this and whether it's pulling a door marked push...

12:07 PM · Jul 25, 2014

Source

<https://twitter.com/smithsam/status/492627127255441408>

Lesson 3

Data will often be
contentious

N-Gram 1:

terrorism

at the

House of Commons



N-Gram 2:

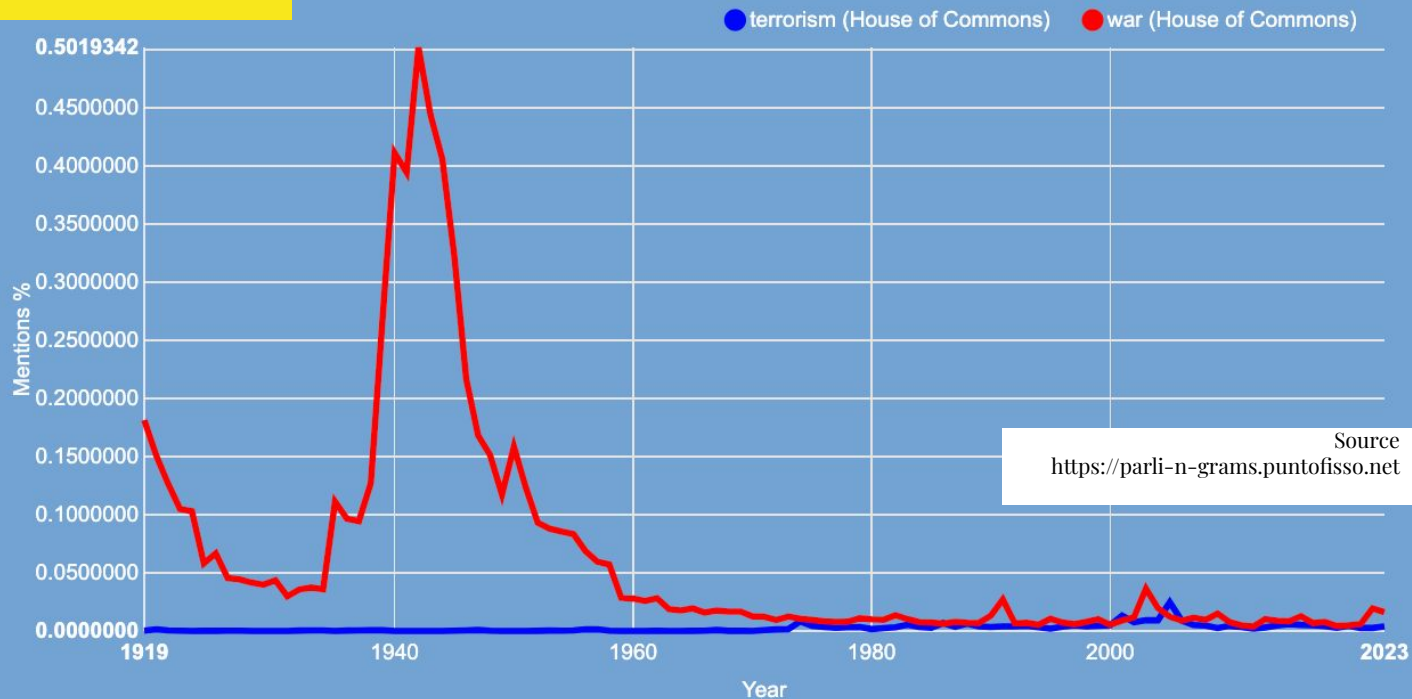
war

at the

House of Commons



Parli-N-Grams



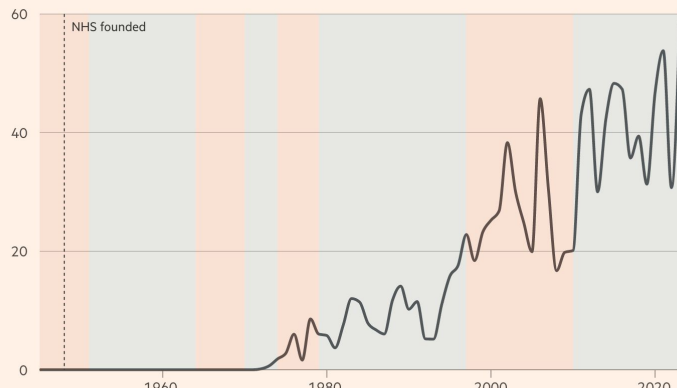
Parli-N-Grams in the media



The NHS has become increasingly politicised

Mentions of 'NHS' per 100,000 words spoken in the House of Commons

Government in power ■ Labour ■ Conservative



A local matter

Mentions of "my constituents" in the House of Commons
as a percentage of all phrases

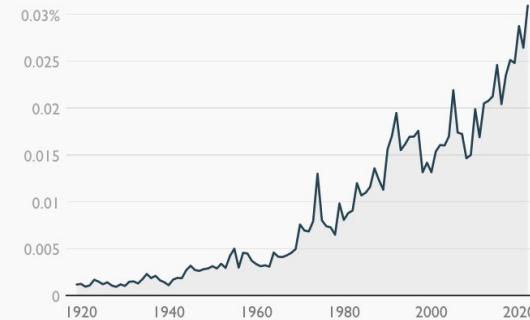


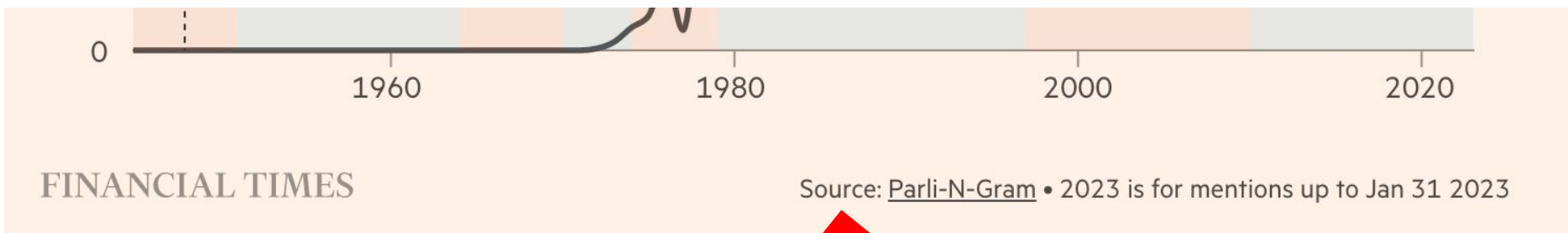
Chart: The Times and The Sunday Times • Source: parli-n-grams.puntofisso.net

Sources

<https://www.ft.com/content/d9db05e7-bb1c-4f38-9a02-bd6b66c9c90b>
Allegra Stratton using Parli-N-Grams on ITV Peston on Sunday in 2018


<https://www.thetimes.co.uk/article/mps-are-becoming-more-local-and-its-making-the-commons-harder-to-control-r5zzpfttc>

Parli-N-Grams in the media



Sources
<https://www.ft.com/content/d9db05e7-bb1c-4f38-9a02-bd6b66c9c99b>
Allegra Stratton using Parli-N-Grams on ITV Peston on Sunday in 2018
<https://www.thetimes.co.uk/article/mps-are-becoming-more-local-and-its-making-the-commons-harder-to-control-r5z2pftre>

Authoritative sources

 UK Parliament

Hansard

House of Commons

UK Parliament > Hansard > Commons Chamber > Oral Answers to Questions > Prime Minister > Engagements

Engagements

Volume 729: debated on Wednesday 8 March 2023

Download text

Previous debate

Next debate

Richard Burgon >
(Leeds East) (Lab)

Q1. If he will list his official engagements for Wednesday 8 March. (903999)

The Prime Minister >
(Rishi Sunak)

Today is International Women's Day. At home, we are taking huge strides to deliver equal opportunities for women, such as mandatory pay gap reporting and the landmark Domestic Abuse Act 2021; and internationally, we have today launched a new women and girls strategy, which puts them at the heart of everything we do.

This morning I had meetings with ministerial colleagues and others. In addition to my duties in this House, I shall have further such meetings later today.

Engagements

Richard Burgon (Leeds East) (Lab)

Q1. If he will list his official engagements for Wednesday 8 March.

The Prime Minister (Rishi Sunak)

Today is International Women's Day. At home, we are taking huge strides to deliver equal opportunities for women, such as mandatory pay gap reporting and the landmark Domestic Abuse Act 2021; and internationally, we have today launched a new women and girls strategy, which puts them at the heart of everything we do.

This morning I had meetings with ministerial colleagues and others. In addition to my duties in this House, I shall have further such meetings later today.

Richard Burgon

Over 100 days ago, the Prime Minister promised to publish his tax returns. He still hasn't. People want transparency in our politics, especially because the Prime Minister is the richest Prime Minister in history and because of the concerns there have been. So why on earth has the Prime Minister not published his tax returns yet, when will he do so, and when he does so, will he include his US tax returns?

The Prime Minister

As I previously confirmed, I will publish my tax returns, and that will be done very shortly.

Anthony Mangnall (Totnes) (Con)

Q4. On International Women's Day, may I congratulate the Prime Minister on all the work that has been done on the preventing sexual violence in conflict initiative and on ensuring that the UK continues to show its global leadership? There are 250,000 people in this country living on park home sites. They are treated as second-class citizens by councils to which they pay council tax, they are treated poorly by managers who own the sites, and they are at a disadvantage when we roll out our

Source

<https://hansard.parliament.uk/commons/2023-03-08/debates/A182F3E8-6CA8-4FBC-B10E-8045D9374E73/Engagements>

DataSets | Additional Datasets

AV Live Logging

Download

⚙️ Learn more

📄 Explore data

Commons Divisions

Download

⚙️ Learn more

📄 Explore data

Commons Oral Question Times

Download

⚙️ Learn more

📄 Explore data

Commons Oral Questions

Download

⚙️ Learn more

📄 Explore data

Commons Written Questions

Download

⚙️ Learn more

📄 Explore data

Election Results

Download

⚙️ Learn more

📄 Explore data

Elections

Download

⚙️ Learn more

📄 Explore data

Lords Bill Amendments

Download

⚙️ Learn more

📄 Explore data

Lords Written Questions

Download

⚙️ Learn more

📄 Explore data

Members

Download

⚙️ Learn more

📄 Explore data

Parliamentary Questions Answered

Download

⚙️ Learn more

📄 Explore data

Publication Logs

Download

⚙️ Learn more

📄 Explore data

Research Briefings

Download

⚙️ Learn more

📄 Explore data

**Authoritative
sources**

Source
<https://explore.data.parliament.uk/>

```

<publicwhip scraperversion="a" latest="yes">
  <major-heading id="uk.org.publicwhip/debate/2023-03-08a.285.0" nospeaker="true" colnum="285"
time="">Prayers - <i>
[Mr Speaker in the Chair]
</i></major-heading>
  <speech id="uk.org.publicwhip/debate/2023-03-08a.285.1" speakername="Lindsay Hoyle" person_id
="uk.org.publicwhip/person/10295" colnum="285" time="" url="">
    <p pid="a285.1/1">Before we come to Women and Equalities questions, may I say how proud I w
as to host an event last night marking The House magazine's publication of its list of 100 wome
n in Westminster? It was an honour to celebrate so many inspiring colleagues. I wish them and e
veryone a happy International Women's Day.</p>
  </speech>
  <oral-heading id="uk.org.publicwhip/debate/2023-03-08a.285.2" nospeaker="true" colnum="285" t
ime="" url="">Oral
Answers to
Questions</oral-heading>
  <major-heading id="uk.org.publicwhip/debate/2023-03-08a.285.3" nospeaker="true" colnum="285"
time="" url="">
Women and Equalities
</major-heading>
  <speech id="uk.org.publicwhip/debate/2023-03-08a.285.4" nospeaker="true" colnum="285" time=""
url="">
    <p pid="a285.4/1">The Minister for Women and Equalities was asked-</p>
  </speech>
  <minor-heading id="uk.org.publicwhip/debate/2023-03-08a.285.5" nospeaker="true" colnum="285"
time="" url="">
Gender Pay Gap
</minor-heading>
  <speech id="uk.org.publicwhip/debate/2023-03-08a.285.6" speakername="Munira Wilson" type="Sta
rt Question" person_id="uk.org.publicwhip/person/25820" colnum="285" time="" url="" oral-qnum="
1">
    <p pid="a285.6/1" qnum="903984">What recent assessment she has made of the implications for
her policies of the gender pay gap.</p>
  </speech>
  <speech id="uk.org.publicwhip/debate/2023-03-08a.285.7" speakername="Maria Caulfield" type="S
tart Answer" person_id="uk.org.publicwhip/person/25397" colnum="285" time="" url="">
    <p pid="a285.7/1">Gender pay gap reporting continues to motivate employers to look at their
pay data and improve workplace gender equality, and huge progress is being made. The gender pa
y gap has fallen by approximately a quarter in the past decade, but of course, there is more wo
rk to be done.</p>
  </speech>

```

Sources that work but
are not authoritative

House of Commons


Thursday 8 September 2022 Meeting started at 9.33am, ended 4.58pm



SHARE 

DOWNLOAD 

INFO 

AUDIO ONLY 

TV logging

AGENDA

INDEX

10:51:08 Rt Hon Penny Mordaunt MP, Lord President of the Council and Leader of the House of Commons (Portsmouth North, Conservative)

10:51:19 Janet Daby MP (Lewisham East, Labour)

10:51:58 Rt Hon Penny Mordaunt MP, Lord President of the Council and Leader of the House of Commons (Portsmouth North, Conservative)

10:52:33 Steve Brine MP (Winchester, Conservative)

10:53:10 Rt Hon Penny Mordaunt MP, Lord President of the Council and Leader of the House of Commons (Portsmouth North, Conservative)

10:53:43 Wera Hobhouse MP (Bath, Liberal Democrat)

Source

<https://parliamentlive.tv/event/index/acb2b28e-2a85-4ac3-a75a-6233a765b37d?in=10:51:08>

start date	2022-09-08T14:20:19Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/140
start date	2022-09-08T09:54:06Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/141
start date	2022-09-08T09:53:43Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/142
start date	2022-09-08T09:53:10Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/143
start date	2022-09-08T09:52:33Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/144
start date	2022-09-08T09:51:58Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/145
start date	2022-09-08T09:51:19Z
- tv clip	http://data.parliament.uk/resources/1506419/videoclips/146
start date	2022-09-08T09:51:08Z

TV logging API

<http://data.parliament.uk/resources/1506419/videoclips/146>

start date	2022-09-08T09:51:08Z
website	<ul style="list-style-type: none"> ◦ https://parliamentlive.tv/event/index/acb2b28e-2a85-4ac3-a75a-6233a765b37d?in=10:51:08 ◦ https://videoplayback.parliamentlive.tv/player/embedcode/acb2b28e-2a85-4ac3-a75a-6233a765b37d?in=2022-09-08T10:51:08&
episode	Main chamber - House of Commons
member	Biography information for Penny Mordaunt

Source

<https://explore.data.parliament.uk/?endpoint=tvprogrammes/1506419>
<https://explore.data.parliament.uk/?endpoint=resources/1506419/videoclips/146>



Petitions

UK Government and Parliament

794

petitions got a response from
the Government

159

petitions were debated in
the House of Commons

Search petitions



[View all open petitions](#)

Popular petitions

Urgently review the current Ofsted Framework and associated methodology

133 signatures in the last hour

Make swift bricks compulsory in new housing to help red-listed birds

113 signatures in the last hour

Abolish Council Tax for State Pensioners

35 signatures in the last hour

Petitions

Source

<https://petition.parliament.uk/>

Using petition ID

<https://petition.parliament.uk/petitions.json>

<https://petition.parliament.uk/petitions/241584.json>

Using resource ID

<http://lda¹.data.parliament.uk/epetitions.json>

<http://lda.data.parliament.uk/epetitions/1075160.json>

<http://lda.data.parliament.uk/epetitions/1075160/track.json>

<http://lda.data.parliament.uk/resources/1075160/governmentresponse.json>

<http://lda.data.parliament.uk/resources/1075160/debate.json>

Signatures by constituencies

<http://lda.data.parliament.uk/epetitions/signaturesbyconstituency/E14000640.json>

<http://lda.data.parliament.uk/resources/872207/signaturesbyconstituency/E14000640>

<http://eldaddp.azurewebsites.net/resources/872207/signaturesbyconstituency/E14000640.json>

Lesson 4

What expectations
should we have on the
availability of
authoritative data?

They Look Like You

Find out which MP you look like the most



Select JPG photo

Choose File

No file chosen

Email

You're not an MP

More parliamentary
data



The average Member of Parliament

The average face of the British Parliament

Source
<https://puntofisso.medium.com/i-calculated-the-average-face-of-a-uk-member-of-parliament-and-heres-what-i-found-37f31b72b5d9>



The average face of the US Congress

US senators respond to 'average' face photo

Source
<https://www.bbc.com/future/article/20171018-this-is-the-face-of-the-average-american-politician>

Lesson 5

As data
wranglers, what
are our ethical
responsibilities?

Data science in a public authority

Source

<http://www.ft.com/cms/s/2/a284331a-9751-11e2-a77c-00144feabdco.html#axzz3XOJKf7LS>

FT Magazine

Home UK World Companies Markets Global Economy Lex Comment
Arts Magazine Food & Drink House & Home Lunch with the FT Style Books Pursuits

Last updated: April 1, 2013 5:35 pm

Why grease is the word in New York



By Gillian Tett

Share Author alerts Print Clip

Comments

Many restaurants don't want to pay disposal costs and fear dealing with organised crime – so they simply tip their grease down a manhole



It is tough to feel cheerful these days about the workings of American government; with the soaring debt and political gridlock, the news from the public sector has been pretty dire. But if you want to see a tiny glimmer of cheer, take a look at a saga that has been unfolding in New York over the issue of “yellow grease”. Yes, you read that right. While politicians in Washington squabble about grandiose fiscal plans, an experiment is quietly under way in the Big Apple about how to handle the hazards posed by the waste from restaurants’ fryers (after they have cooked chips, burgers, prawn rolls and so on).

And though this so-called yellow grease problem is not something that Financial Times readers probably wish to ponder – least of all over breakfast – the tale is oddly inspirational. For it shows how cash-strapped governments can make resources

NaPTAN

NaPTAN

National

NaPTAN

National Public

NaPTAN

National Public

Transport

NaPTAN

National Public

Transport Access

NaPTAN

National Public

Transport Access Node

NaPTAN

National Public

Transport Access Node

BETAThis is a new service – your [feedback](#) will help us to improve it.**NaPTAN**

NaPTAN and NPTG

Home to the National Public Transport Access Nodes (NaPTAN) and the National Public Transport Gazetteer (NPTG) datasets

Welcome

Use this service to access, edit and work with the NaPTAN and NPTG datasets.

These datasets:

- help locate transport access points
- include bus, rail, tram, metro, underground, air and ferry services
- include locality information

NaPTAN

15 years

500,000+ records

386 local authorities

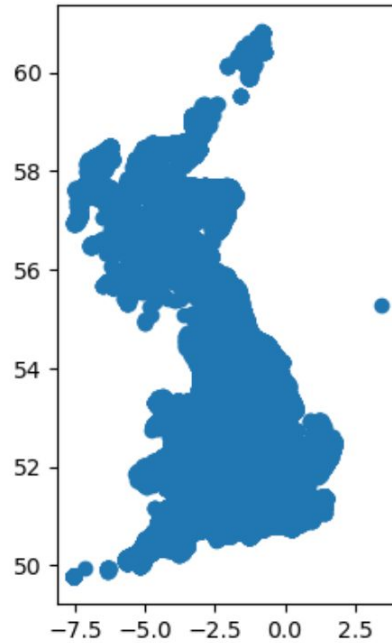
Legally required

NaPTAN

```
ATCOCode,NaptanCode,PlateCode,CleardownCode,CommonName,CommonNameLang,ShortCommonName,ShortCommonNameLang,Landmark,LandmarkLang,Street,StreetLang,Crossing,CrossingLang,Indicator,IndicatorLang,Bearing,NptgLocalityCode,LocalityName,ParentLocalityName,GrandParentLocalityName,Town,TownLang,Suburb,SuburbLang,LocalityCentre,GridType,Easting,Northing,Longitude,Latitude,StopType,BusStopType,TimingStatus,DefaultWaitTime,Notes,NotesLang,AdministrativeAreaCode,CreationDateTime,ModificationDateTime,RevisionNumber,Modification,Status
0100BRP90310,bstgwpa,,Temple Meads Stn,,Temple Meads Stn,,,Redcliffe Way,,,T3,,SE,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359396,172388,-2.58569,51.44901,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2019-09-13T10:41:24,92,new,active
0100BRP90311,bstgwpm,,Temple Meads Stn,,Temple Meads Stn,,,Victoria Street,,,T7,,NW,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359403,172513,-2.5856,51.45014,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2019-09-13T10:41:38,73,new,active
010000056,bstpjpgw,,Temple Meads Stn,,Temple Meads Stn,,,Temple Gate,,,T2,,N,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359452,172461,-2.58489,51.44967,BCT,MKD,OTH,,,009,2019-08-06T00:00:00,2019-08-06T11:09:11,120,new,active
0100BRP90312,bstjaja,,Temple Meads Stn,,Temple Meads Stn,,,Station Approach,,,T6,,NE,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359609,172383,-2.58262,51.44898,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2020-11-26T13:23:29,45,new,active
0100BRP90313,bstjama,,Temple Meads Stn,,Temple Meads Stn,,,Station Approach,,,T5,,NE,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359669,172433,-2.58177,51.44944,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2020-11-26T13:22:13,44,new,active
0100BRP90317,bstjada,,Temple Meads Stn,,Temple Meads Stn,,,Temple Gate,,,T1,,SE,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359581,172304,-2.58302,51.44827,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2019-08-06T11:10:46,49,new,active
0100BRP90318,bstgwtd,,Temple Meads Stn,,Temple Meads Stn,,,Redcliffe Way,,,T4,,NW,N0077020,Temple Meads,Bristol,,,,,false,UKOS,359439,172371,-2.58507,51.44886,BCT,MKD,OTH,,,009,2009-08-25T00:00:00,2019-08-06T11:11:05,86,new,active
0100BRA46646,bstpdpm,,Old Market Str,,,,,Old Market Street,,,M3,,E,N0077040,Old Market,Bristol,,,,,false,UKOS,359635,173156,-2.58234,51.45593,BCT,MKD,OTH,,,009,2007-03-14T00:00:00,2018-07-18T13:43:01,23,new,active
0100BRA46647,bstpdpt,,Old Market Str,,,,,Old Market Street,,,M4,,W,N0077040,Old Market,Bristol,,,,,false,UKOS,359706,173174,-2.58132,51.4561,BCT,MKD,OTH,,,009,2007-03-14T00:00:00,2018-06-11T14:45:46,17,new,active
```

```
gdf.plot()
```

<Axes: >

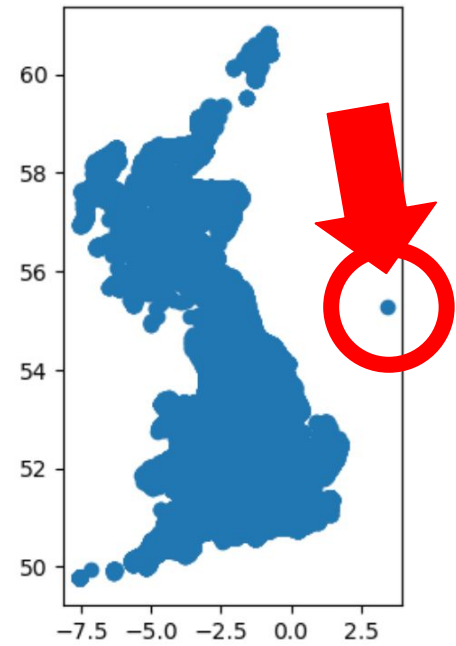


NaPTAN

Source
<https://github.com/puntofisso/csv-conf-v7>

```
gdf.plot()
```

<Axes: >



NaPTAN

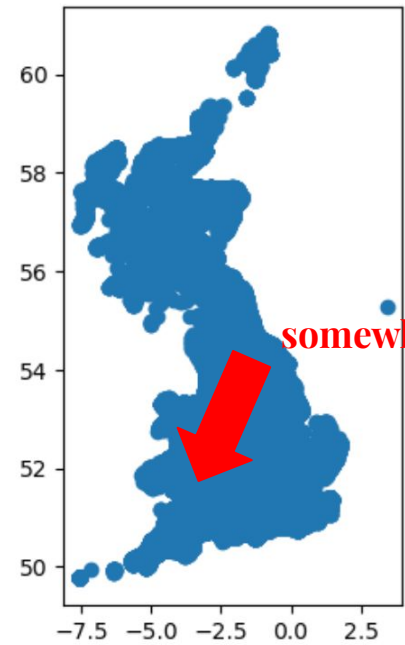
Source
<https://github.com/puntofisso/csv-conf-v7>

```
# Display suspicious point
gdf[ (gdf['Longitude'] > 2.5)]
```

ATCOCode	NaptanCode	CommonName	ShortCommonName	Landmark	Street	Indicator	Bearing	NptgLocalityCode	LocalityName	Pa
375623	5710WDB48478	cditpgj	Penarth Heights	NaN	NaN	Arcot Street	not in use	N	E0035848	Penarth Moors

```
gdf.plot()
```

<Axes: >



NaPTAN

Source
<https://github.com/puntofisso/csv-conf-v7>

```
# Display suspicious point
gdf[ (gdf['Longitude'] > 2.5)]
```

	ATCOCode	NaptanCode	CommonName	ShortCommonName	Landmark	Street	Indicator	Bearing	NptgLocalityCode	LocalityName	Pa
375623	5710WDB48478	cditpgj	Penarth Heights	NaN	NaN	Arcot Street	not in use	N	E0035848	Penarth Moors	



NaPTAN Error Log

NaPTAN

NaPTAN Error Log

Please list all the errors you find on the ground as you survey, where the NaPTAN data does not match reality. This does not include positional errors which should just be corrected as you edit.

Adding any photos would be useful too.

	ATCO Code	Location	Error	User	Corrected
1	43002256501	The College Arms Shaftmoor Lane	NaPTAN Indicator says Stop SD. Actual bus stop is labelled Springfield SC	--Brianboru 15:29, 6 April 2009 (UTC)	
2	43002102502	Stratford Rd Springfield Rd	NaPTAN Indicator says Stop SC. Actual bus stop is labelled Springfield SF, which duplicates the correct labelling of 43000225604	--Brianboru 15:33, 6 April 2009 (UTC)	
3	43000955922	West Brom Bus Station	This bus stop which NaPTAN labels as Stand X no longer exists	--Brianboru 15:43, 6 April 2009 (UTC)	
4	43000955921	West Brom Bus Station	This bus stop which NaPTAN labels as Stand W no longer exists	--Brianboru 15:43, 6 April 2009 (UTC)	
5	43000955923	West Brom Bus Station	This bus stop which NapTAN labels as Coach Stand I've assumed is allocated to Stand O which has no route information displayed unlike all the other bus stands in the bus station	--Brianboru 15:43, 6 April 2009 (UTC)	
6	43000445202	Swan Shopping Centre Yardley	No evidence for this bus stop on the ground. Resurveyed, replaced by temporary stop sign but the road is marked with a bus stop bay	--Brianboru 16:22, 6 April 2009 (UTC) --Brianboru 18:51, 22 May 2009 (UTC)	

Source

https://wiki.openstreetmap.org/wiki/NaPTAN_Error_Log

Main Page
The map
Map Features
Contributors
Help
Blogs
Shop
Donations
Wiki discussion
Recent changes

Tools

What links here
Related changes
Special pages
Printable version
Permanent link
Page information
Cite this page

NaPTAN

```
# Using Lon/Lat
gdftouse = gdfGlasgow
graph = Glasgow_proj

for row in range(len(gdftouse)):

    lon = gdftouse.at[row, 'Longitude']
    lat = gdftouse.at[row, 'Latitude']

    point = lat, lon

    G = ox.graph_from_point(point, network_type='drive')
    Gp = ox.project_graph(G)
    point_geom_proj, crs = ox.projection.project_geometry(Point(reversed(point)), to_crs=Gp.graph['crs'])

    x, y = point_geom_proj.x, point_geom_proj.y
    nearest = ox.nearest_edges(Gp, x, y, return_dist=True)

    print(gdftouse.at[row, 'ATCOCode'] + " " + str(nearest))
```

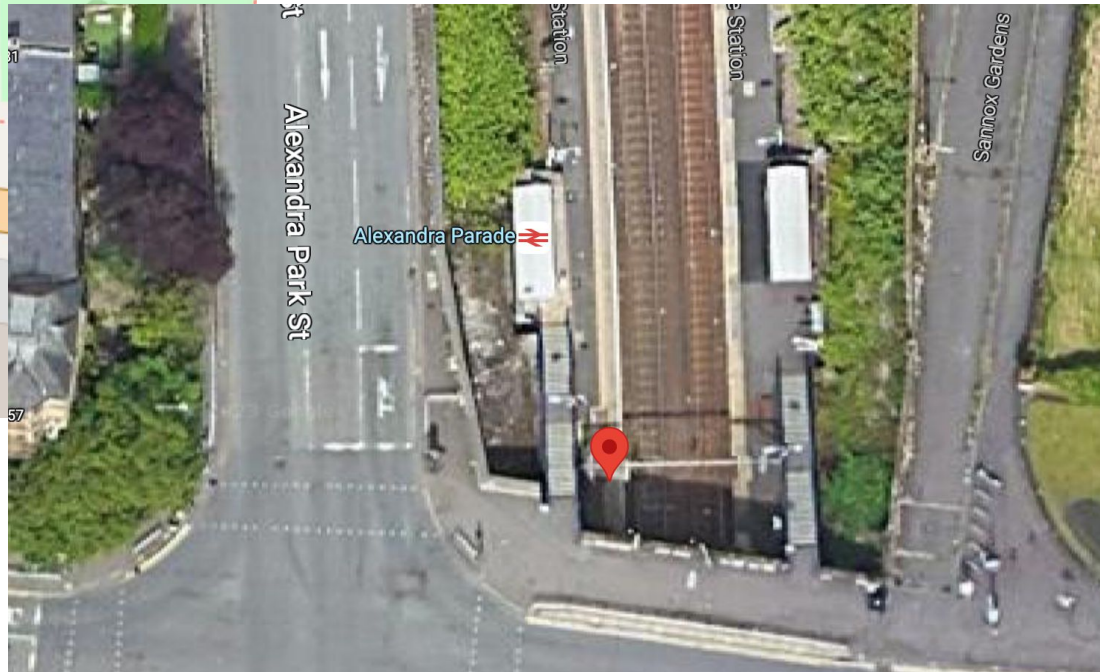
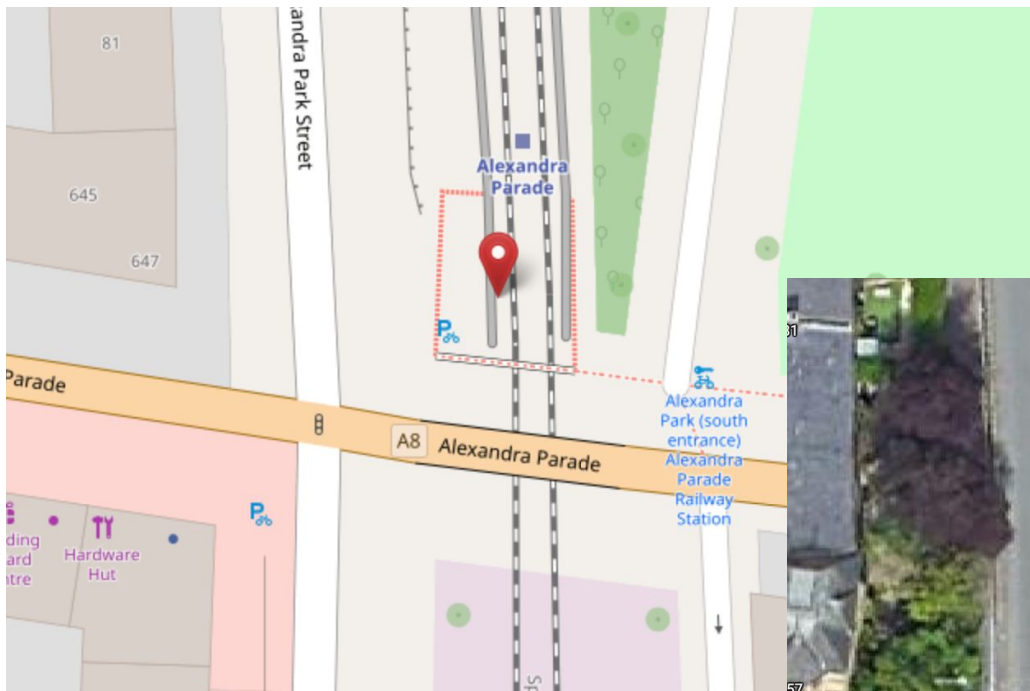
```
9100ALXPD ((27432018, 27432022, 0), 15.510528195359685)
9100ANSL ((86873440, 289767929, 0), 32.86422400972587)
9100ASFIELD ((263679063, 263679059, 0), 10.600554772698116)
9100BALISTN ((749883364, 778425769, 0), 37.11927753808497)
9100BLGRVE ((27437069, 28096025, 0), 55.71026933270662)
9100BNHL ((338716228, 295026202, 0), 23.121385244014373)
9100BRIDGTN ((30039549, 94610187, 0), 14.6442920784617)
9100CARMYLE ((289895757, 2096433672, 0), 75.23439203465473)
9100CCRT ((986404011, 8763806081, 0), 34.1202391305199)
9100CDNL ((77895651, 77895656, 0), 9.948066817973286)
9100CFFT ((262841270, 262841518, 0), 59.653235513847314)
9100CKHL ((6605901604, 60676358, 0), 46.92661116164154)
9100CNTY ((581076901, 270384239, 0), 29.692528874358064)
9100CROKSTN ((33725705, 29107370, 0), 27.87283952901534)
9100CROSSHL ((450849971, 450849967, 0), 26.94564590486191)
9100CRSMYLF ((28783706, 2202399156, 0), 29.754153016205457)
9100DAWSHLM ((2913998, 60038020, 0), 23.64201286277117)
```



distance in metres

Source
<https://github.com/puntofisso/csv-conf-v7>

NaPTAN







Source
Google Maps and Open Street Map for the coordinates
55.86317091519,-4.21064800754

Bus stations







Image CC BY-SA Giuseppe Sollazzo

 Buses
Stop  Towards
Harringay
 Tottenham or Stamford Hill

 High Road
(east side)
Green Lanes

 Way out 

Westbury 
Avenue

 Buses
Stop  Towards
Muswell Hill
Palmers Green or Enfield
Great Cambridge Road and Edmonton



Bus stations

BUS 
STATION

“Can you give me the
list of all **bus stations** in
the **UK?**”

Lesson 6

Datasets should evolve in line with their purpose; there is sometimes a policy/data disconnect

Data » AI.

How to do
it well?

**“YOU CAN’T EXPLAIN
HOW AI WORKS!”**

“AI WILL TAKE MY JOB!”

THE
SIMPLE ANSWERS
TO THE QUESTIONS THAT GET ASKED
ABOUT EVERY NEW TECHNOLOGY:

WILL <input type="checkbox"/> MAKE US ALL GENIUSES?	NO
WILL <input type="checkbox"/> MAKE US ALL MORONS?	NO
WILL <input type="checkbox"/> DESTROY WHOLE INDUSTRIES?	YES
WILL <input type="checkbox"/> MAKE US MORE EMPATHETIC?	NO
WILL <input type="checkbox"/> MAKE US LESS CARING?	NO
WILL TEENS USE <input type="checkbox"/> FOR SEX?	YES
WERE THEY GOING TO HAVE SEX ANYWAY?	YES
WILL <input type="checkbox"/> DESTROY MUSIC?	NO
WILL <input type="checkbox"/> DESTROY ART?	NO
BUT CAN'T WE GO BACK TO A TIME WHEN—	NO
WILL <input type="checkbox"/> BRING ABOUT WORLD PEACE?	NO
WILL <input type="checkbox"/> CAUSE WIDESPREAD ALIENATION BY CREATING A WORLD OF EMPTY EXPERIENCES?	WE WERE ALREADY ALIENATED

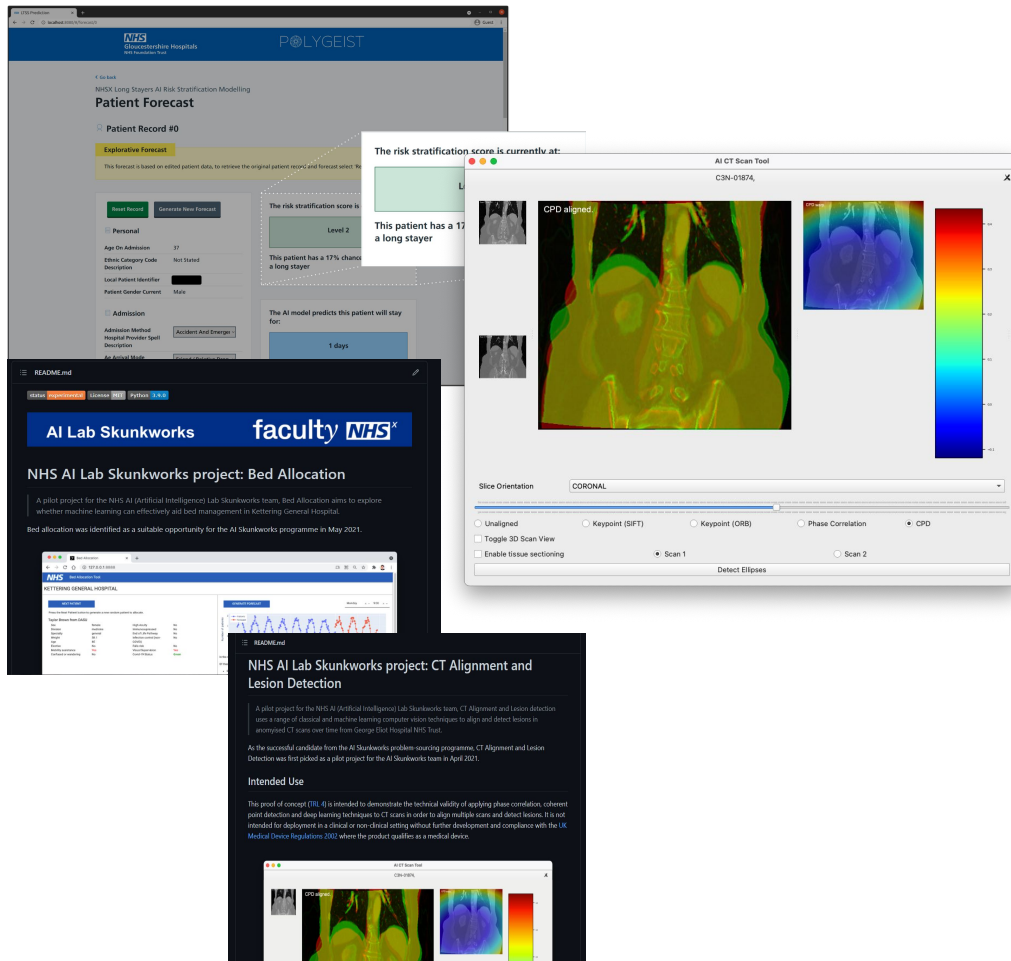
XKCD on “every new
technology”

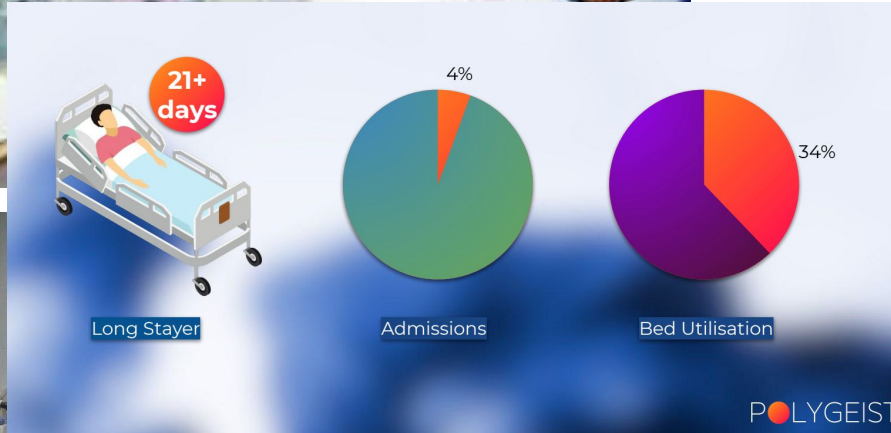
Source
<https://xkcd.com/1289/>

Open Source is a core contributor to understanding and challenging a new tech

Source

<https://nhsx.github.io/skunkworks/>
<https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/skunkworks/>



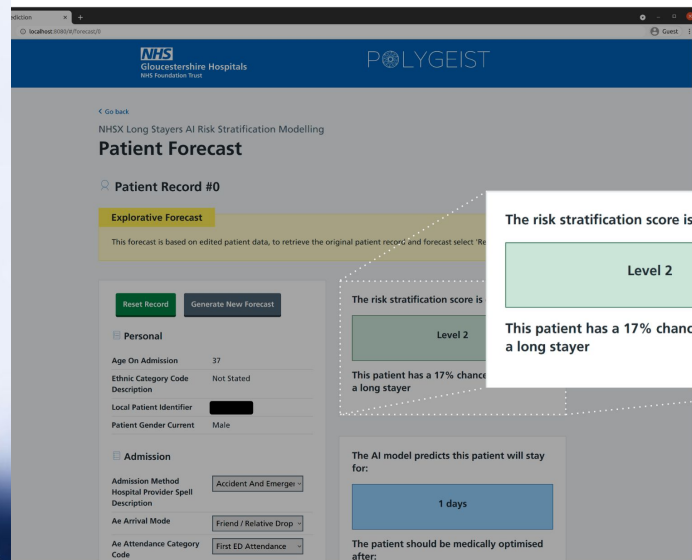


Predicting a patient's length of stay in hospital



Source
<https://github.com/nhsx/skunkworks-long-stayer-risk-stratification>

Pipeline of decision making.



<https://github.com/nhsx/skunkworks-long-stayer-risk-stratification>

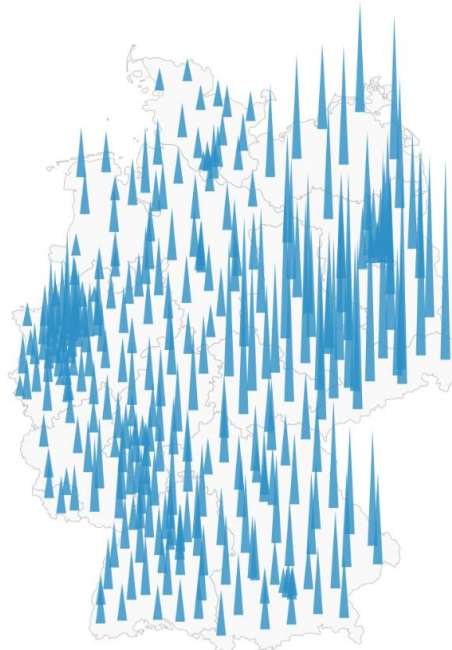
“What do we
exactly *do* with
this prediction?”

Lesson 7

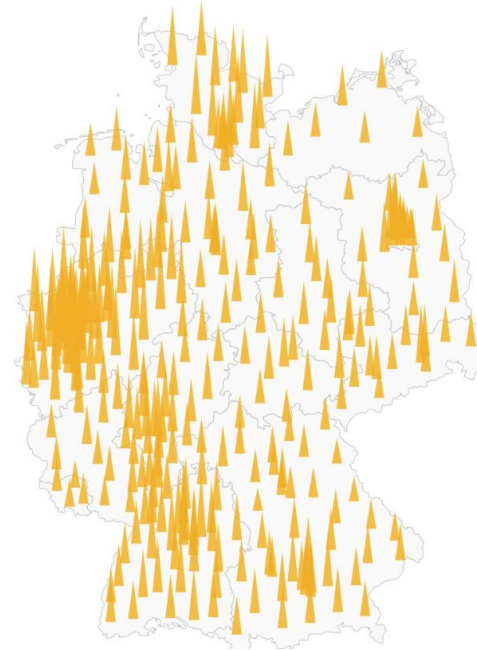
Managing a data-driven service or product is not the same as managing any other IT service

*Deutschland ist in allen 299 Wahlkreisen nach rechts
gerückt*

New dataviz approaches



AfD



Source

<https://interaktiv.morgenpost.de/analyse-bundestagswahl-2017/>

Replicable data journalism

Small multiples for the German elections 2017

Raw data and methodology

Marie-Louise Timcke, Interaktiv-Team Berliner Morgenpost

03.10.2017

For our [analysis of the German election data](#), we prepared a script in the statistical programming environment R. The script was divided in two main parts: A preprocessing part in which we cleaned and structured the raw election data as we got it directly from the electoral management body, and the analysis part that transformed the election data into tables, models and maps.

All steps in the preprocessing were made to fit the data from the 2013 elections and were adjusted for the 2017 data when the first extrapolations were published. In this Notebook we show how we structured the original dataset in order to work with it and how we created the small multiples published in our project.

Preprocessing the data from the electoral management body

```
# install.packages("needs");library(needs)
needs(tidyverse, magrittr, rgdal, grid, gridExtra, mapproj, rgeos) # loading needed packages
# tidyverse, magrittr, stringr: to manipulate and structure data
# rgdal, mapproj, rgeos: to load geospatial data, apply map projections and calculate shape centroids
# grid, gridExtra: to arrange the maps in a grid for this Notebook
```

Load the election data

```
d <- read.csv("https://www.bundeswahlleiter.de/dam/jcr/72f186bb-aa56-47d3-b24c-6a46f5de22d0/btw17_kerg.csv", header=F, sep=";", skip=1, stringsAsFactors = F, na.strings="")

head(d[1:9])
```

##

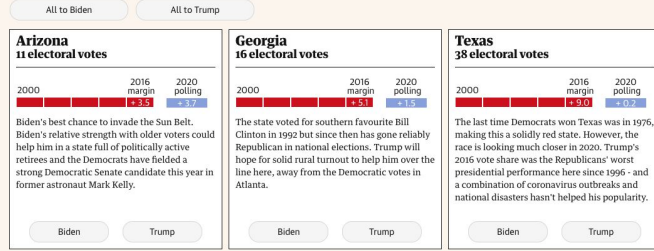
v1

Source
https://interaktiv.morgenpost.de/analyse-bundestagswahl-2017/data/btw17_analysis.html



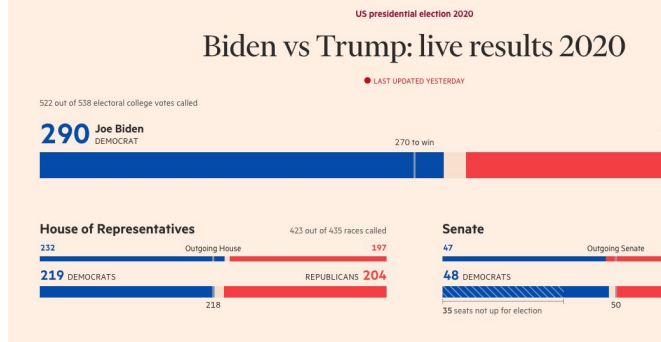
Formerly solid Republican states

Usually solid Republican territory, these reddish states look like they could be in play for Biden in 2020. If he can win one of the three, it could be curtains for Trump.

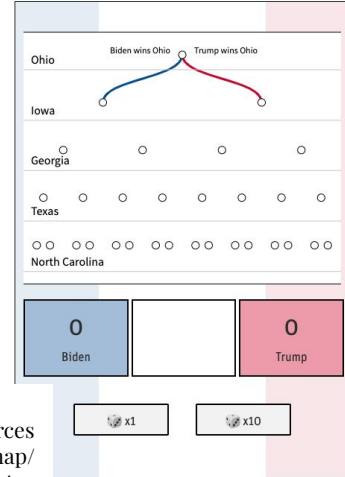
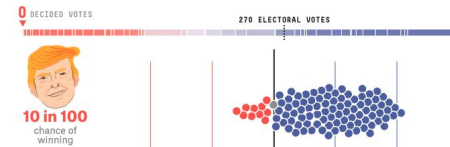
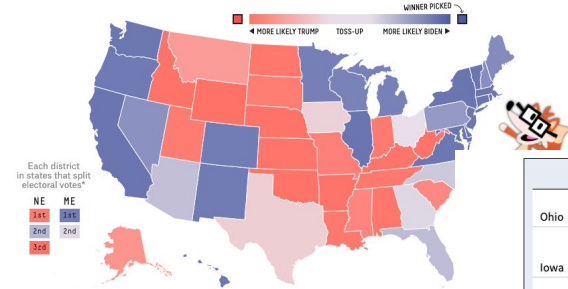


Important swing states

These states have been trending Republican in recent presidential elections, but Biden could put them back in play for the Democrats in 2020.



FiveThirtyEight



Predictive data journalism

Sources

<https://projects.fivethirtyeight.com/trump-biden-election-map/>

<https://ig.ft.com/us-election-2020/#calculator>

<https://www.newstatesman.com/international/2020/11/us-2020-presidential-election-forecast-model-will-donald-trump-or-joe-biden>

<https://www.theguardian.com/us-news/ng-interactive/2020/oct/30/build-your-own-us-election-result-plot-a-win-for-biden-or-trump>

Since when is it our job to predict?

Mona Chalabi, Guardian US Data Editor
Today in Focus, 22 October 2020

Source

<https://www.theguardian.com/news/audio/2020/oct/22/us-election-2020-can-we-trust-the-polls-podcast>

Ost-West? Nord-Süd? Oder ganz anders?

Teile Deutschland und entdecke die Unterschiede

30 Jahre nach der Wiedervereinigung bleibt Deutschland in vielen Belangen ein geteiltes Land. Doch die Grenzen verlaufen nicht nur zwischen Ost und West.

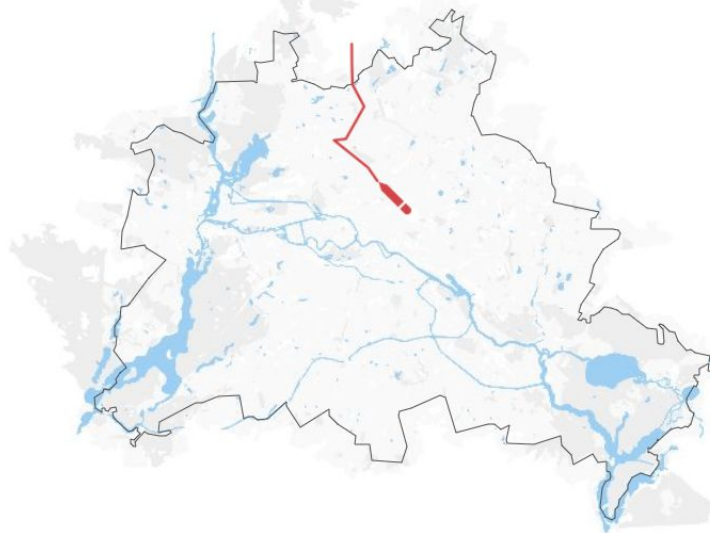
Zerschneide die Deutschlandkarte in zwei beliebige Teile und entdecke die Unterschiede bei 30 Merkmalen - von Millionärsdichte bis Graffiti!



Wissen Sie noch, wo die Mauer Berlin teilte?

Genau 10.315 Tage trennte die Mauer die Berliner, 43 Kilometer verliefen direkt durch das Stadtgebiet. Können Sie zeigen, wo das war?

Zeichnen Sie den Verlauf der Berliner Mauer ein

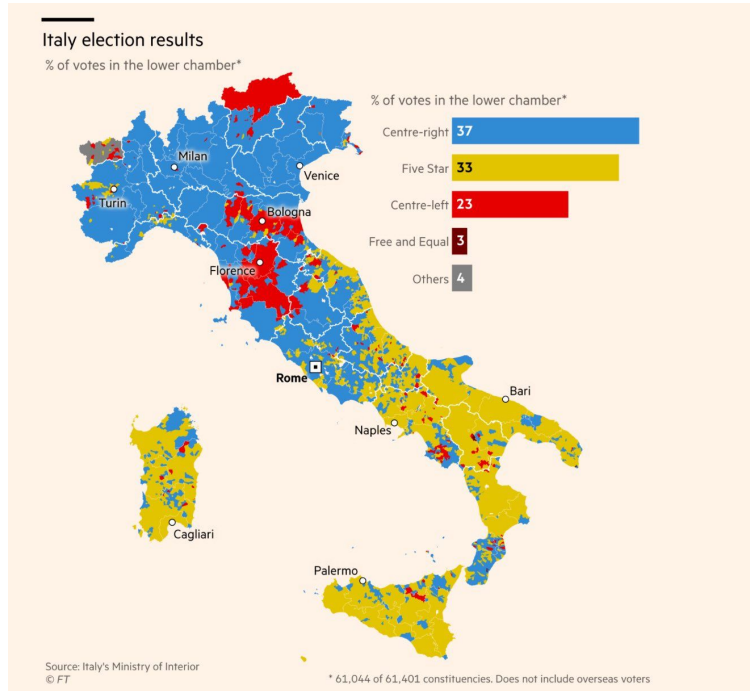


Interactive data journalism

Lesson 8

Data maturity
and data literacy
are improving

Two maps, three (?) Italies

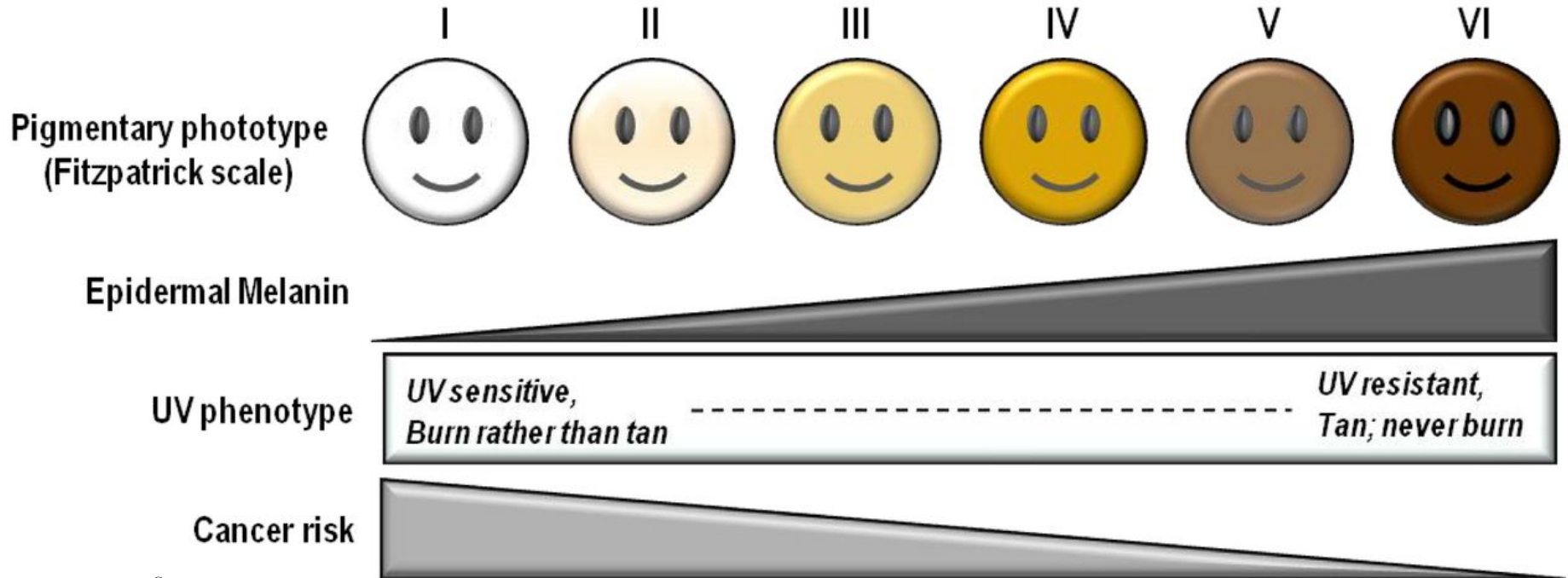


Sources
<https://www.ft.com/content/d1902f6-2062-11e8-a895-1ba1f72c2c11?segmentId=6132a895-e068-7ddc-4cec-a1abfa5c8378>
<https://medium.com/@puntofisso/le-vere-mappe-delle-elezioni-italiane-a0cb89d27d9e>

Lesson 9

Data-driven does not
mean impartial

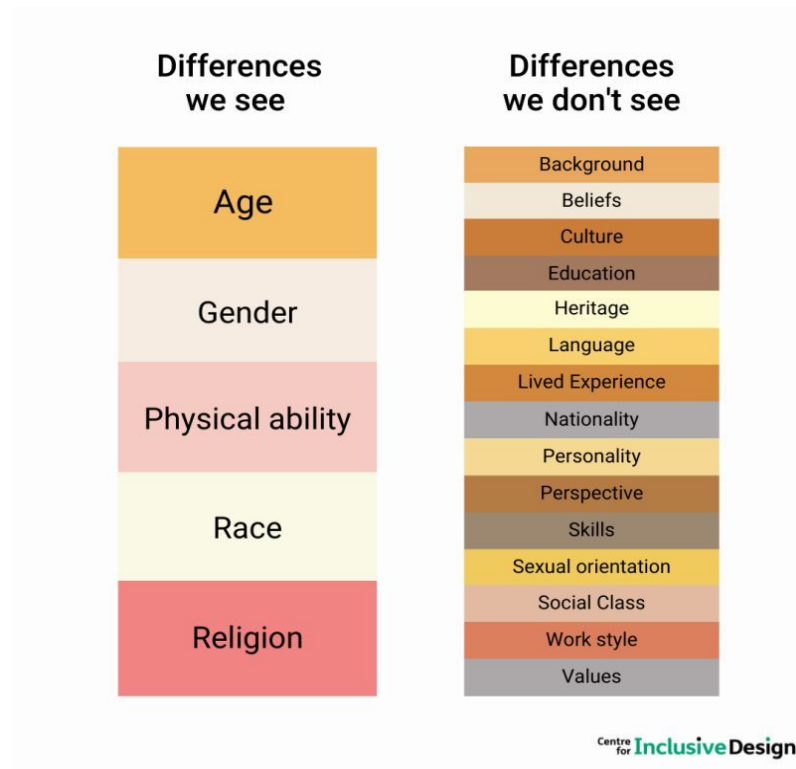
Bias and AI



Sources

https://en.wikipedia.org/wiki/Fitzpatrick_scale

Bias and AI



Source
<https://www.linkedin.com/feed/update/urn:li:activity:6953475768391782400/>

Data-driven harm

Exploring social justice
in an age of datafication

Data Harm Record

Data Harm Record (Updated)

Updated August 2020

Joanna Redden, Jessica Brand and
Terzieva

Source

<https://datajusticelab.org/publications/>

<https://incidentdatabase.ai/>

AIID

AI INCIDENT DATABASE

English ▾

Twitter Facebook LinkedIn

Discover

Submit

Welcome to the AIID

Discover Incidents

Spatial View

Table View

Entities

Taxonomies

Word Counts

Submit Incident Reports

Submission Leaderboard

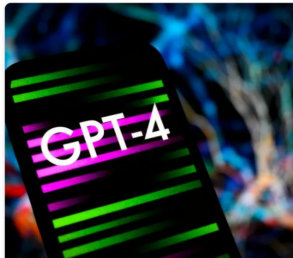
Blog

Subscribe

Welcome to the
AI Incident Database

Search

Discover



Latest Incident Report

The ChatGPT bug exposed more private data than previously thought, OpenAI confirms

2023-04-18

A ChatGPT bug found earlier this week also revealed user's payment information, says OpenAI.

The AI chatbot was shut down on March 20, due to a bug that exposed titles and the first message of new conversations from active users' chat histo...

Read More →

“Applying **intersectionality** to data and data science is useful to understand how injustice can be embedded in the administration of power.”

Donata Columbro

Author, journalist, and data-humaniser

“Dentro l’algoritmo”, Effequ, 2022

Lesson 10

Data can be
harmful

The “data release target” at the UK DEFRA (2015)

Vast data reserves from Defra are set to transform the world of food and farming in the single biggest government data giveaway the UK has ever seen, Environment Secretary Elizabeth Truss will announce today.

Outlining her vision for the future of British food, farming and the natural environment, the Environment Secretary will say that over the next year, virtually all the data Defra holds—at least 8,000 sets—will be made freely available to the public, putting Britain at the forefront of the data revolution.

Harnessing this open data will help food and farming achieve its full potential and enable the UK to become a one-nation economy, where the productivity of the countryside will be brought up to the level of our towns and cities.

Source

<https://www.gov.uk/government/news/environment-secretary-unveils-vision-for-open-data-to-transform-food-and-farming>

The McKinsey report on Open Data

Open data—public information and shared data from private sources—can help create \$3 trillion a year of value in seven areas of the global economy.

Source

<http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>

A special type of hype: snake oil

Incomplete & crude but useful breakdown

Genuine, rapid progress

- Shazam, reverse img search
- Face recognition
- Med. diagnosis from scans
- Speech to text
- Deepfakes

Perception

Imperfect but improving

- Spam detection
- Copyright violation
- Automated essay grading
- Hate speech detection
- Content recommendation

Automating
judgment

Fundamentally dubious

- Predicting recidivism
- Predicting job success
- Predictive policing
- Predicting terrorist risk
- Predicting at-risk kids

Predicting
social outcomes

Lesson 11

There is hype
about data

“Critical friends”, part one

Replying to [@puntofisso](#) [@NHSTransform](#)

You have crisis in primary care, GP suicides, chaos in secondary care, social care fail, food price hikes, fuel price hikes, and a total economic collapse in the horizon.

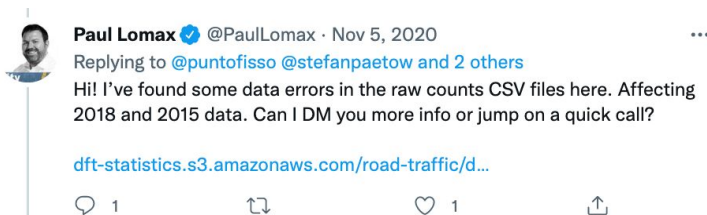
Absolutely, this is the time to ask stupid questions about what [#HASHTAGai](#) can solve.

Sack yourselves. Now

10:46pm · 8 Aug 2022 · Twitter for Android

“Critical friends”, part two

Nov 2020



Oct 2022



Open data as the conduit to good things



“Opening our data was at most half the issue. More important was **the effect it had on open working** among colleagues across many organisations in the DEFRA group, and with a range of external partners from universities to charities to commercial organisations.

That led to interesting discussions about **how open government relates to accountability**, and also to user-centred design.

It took us into how we can be more open as individuals, sharing a bit more about ourselves as a way of making connections with others: building respect and trust.

Open has all sorts of dimensions. It encompasses being open to challenge, open to new experiences, open to new ideas and to difference”.

– Clare Moriarty

(then) Permanent Secretary at the UK Department for Environment, Food & Rural Affairs

Source

<https://www.fda.org.uk/home/Newsandmedia/Features/clare-moriarty-women-into-leadership-speech.aspx>

Lesson 12

Open data and
open source are
transformational

Talking with data

Multidisciplinary, Holistic

Evolving, Iterative

Open, Engaging, Intersectional

Consequence-aware, not impartial



Thank you!

Talking with data

Stories and lessons from my data adventures

Giuseppe Sollazzo • @puntofisso • csv,conf,v7
<http://puntofisso.net/csvconf2023>