SO DU 2023

The 4th Scottish Open Data Unconference

Aberdeen 18-19 November 2023

The unconference was organised and hosted by Code the City (charity SC047835). We are grateful to sponsors The Scottish Government's Data Division and DAMA UK, and to venue operators, One Tech Hub, without whose generosity the event could not take place.
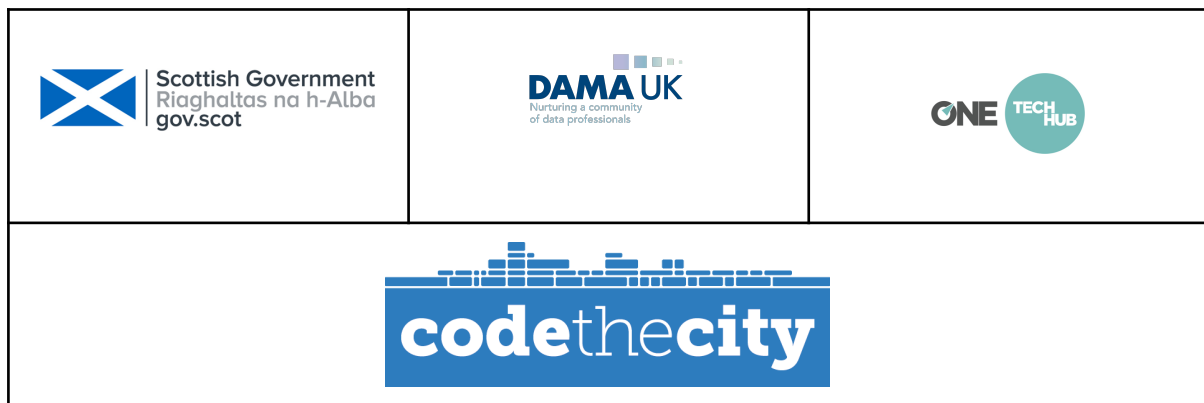
But most of all we are grateful to the participants from across the UK: from academia, industry, government and civic society, who came together over a weekend to share, explore, explain, discuss, and dream.
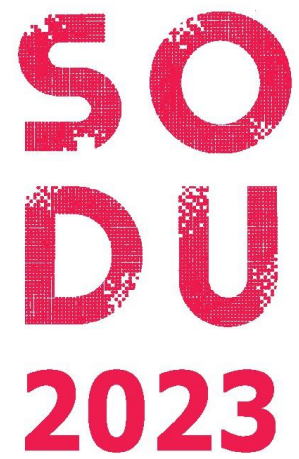
The following pages are notes captured live during the 30 or so sessions which participants proposed and ran.

As we continue the conversation and collaboration, you can keep up with it all on CTC's Open Data Scotland Slack group.

See you all again in 2024!

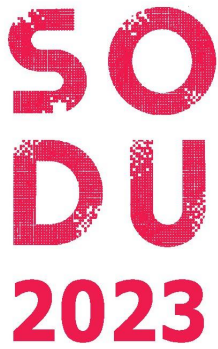Ian, Steve, Karen, Jack, Arrash and Bruce
Code the City

SODU 2023

Day One

18 November 2023

SODU 2023

Day 1 - Session 1 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- [Find.data.gov](#).scot

Leader(s)

- Name: Gautham and Shona Nicol

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Running projects around

- Data maturity - cohort running now and a new one in January - concept to get people to think differently about their data assets and become more data literate.
- Set up a community of practice to help public sector organisations to improve standards in a FAIR way

Want to understand how we are using Open Data, looking for use cases and feedback on the data we are finding.

Question - Is this to replace statistics.gov - answer is no, its not a portal, it stores some meta data but does not store the data. Acknowledges they have a lot of portals and aim is to make it more discoverable - its more an API that can be embedded in other data portals which say nature.scot

Schema.org is sat behind it and it's hoped that Google will bring it to the forefront.
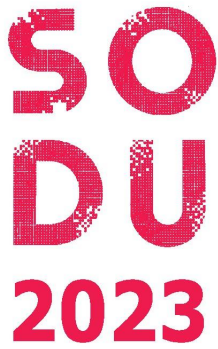
Find.data.gov.scot is the Scottish version of Dtechtive.

- Existing data search engines only discover data sets in the surface web published using standards and frameworks
- If a website doesn't follow schema org unlikely to appear in Google dataset search
- Big gap of around 90%
- Dtechtive is a data search engine that discovers datasets others cannot reach
- This is made possible by its AI powered bottom up approach to dataset creation and acquisition
- Why aren't datasets found?
  - Websites failing to apply standards
  - Data tables in HTML
  - Date inside PDF
  - Datasets require login access
  - Datasets behind a paywall

Covid19 example if you search on google datasets for 'Covid19 scotland' would get 69 datasets, if you do it on Dtechtive then there are 84

Dtechtive indexes the metadata - does not store any data.

A new release will see Dtechtive generating csvs

SODU 2023

Day 1 - Session 1 - The Lab

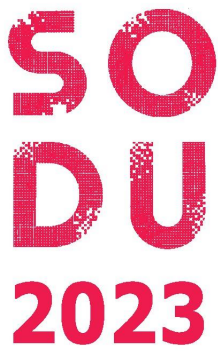You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Leader: Jez

OpenStreetMap / Overture maps - What?

- TomTom Overture maps https://overturemaps.org/
- Licensing issues?
- https://osmuk.org/
- How do you build an OpenStreetMap local community?
- Humanitarian
  - https://www.missingmaps.org/
  - https://www.hotosm.org/projects/missing_maps
  - https://www.hotosm.org/
- Civic
  - https://openwastemap.uk/
  - https://healthsites.io/
  - https://maproulette.org/
- Is the OpenStreetMap ecosystem "working"?
  - Community building? - Is there a platform?
  - How do we connect mappers in the same area? "Find mappers near me"
  - How do we make the platform more accessible to people?
    - The terminology is "dense"
  - "Community of purpose"
- Turf Game - https://turfgame.com/
  - Pokemon Go-esque gathering of information
  - Can we harness information from existing apps?

Leader(s)

- Name: Jez Nicholson

1

SODU 2023

Day 1 - Session 2 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Connecting Open Data to LLMs / Machine Learnjng
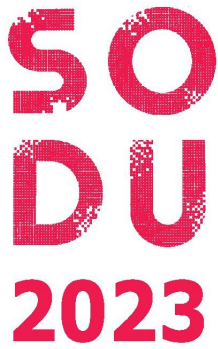
Leader(s)

- Name: Steve Shearman

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- Code way for connecting information with computers
- Easier interface between human instruction and computers
- How to get the best out of ChatGPT?
- Logic of coding another more rigid language
- Human languages evolve and rules get broken
- Changes in programming languages not so fluid
- Requirements need to be clear in business
- Generally requires iteration
- If this goes wrong, you won't develop what you want
- ChatGPT has infinite interns
- May have to iterate a few times
- LLMs trying to simulate human chatting
- They are an interpolation of data
- LLMs can't do research
- Giving LLMs a concise set of requirements can give you something back
- Can hallucinate results

- If given an error, can copy and paste back in and can improve on this
- How to communicate the best way to a LLM
- Can push off responsibility to other people
- Should spend time to try out ChatGPT
- Other humans aren't perfect either
- Good to find reliable data for training and allow machine learning take context into its training
- Detective – dataset. Google main search engine. Aggregating large datasets in Scotland. Building LLMs to use the datasets. Developing datasets that people can use and transform the datasets into a csv file so its ready to use. Can simplify data analysts job
- Businesses may use ChatGPT
- Way to make the API for ChatGPT better
- Need to figure out what is valuable to you
- How to get data?
- How to take data and give it to LLM?
- Don't need to be a coder to write code as now have an interface
- Knowing where the data is and whether the data is relevant
- There is lots of data and assuming we can get it, how do we get the interface?
- With LLMs it is hard to get the results from them that you want
- Need the correct input prompt
- Best way to ask is if I was asking a human to do this, what would I ask?
- Can use one chat to build a better chat
- Language models can infer data so only need to give them the first few samples
- System prompt -> giving the language model a set up role to write in
- Can use ChatGPT to rewrite code
- Can use to suggest which parts of code can be improved
- Can repeat errors - it has a memory of a maximum amount of data it can remember (context window). When the chat goes beyonds this, it then can start repeating stuff and forgets the earlier information
- This problem is reduced as newer language models have a larger memory
- There are techniques where a database build with the data required and when asking a question it pulls out the data it needs
- Can develop language models (ChatGPTs) specific for the task you want
- Is there a specific way to find tune a LLM?
- Good niche cases when have databases with you
- How to use ChatGPT to analyse a large data set (like the NHS dataset)
- Could go and collect data and then return a summary of that dataset
- Need to write a function to collect that data and then ChatGPT can use it to collect this data (ChatGPT extension)
- Good to have community to playtest ChatGPT
- Interface -> instead of human, coder, computer now have human, language model, coder, computer
- Companies are now switching to using Copilot
- Can ask language model how to test code as well
- Testing -> if have context and expected answers can then test it
- Can tell ChatGPT to go and run the tests
- What is the commercial model behind ChatGPT?
- Can write a frontend which interacts with ChatGPT on the backend
- Would have to pay for the calls to the ChatGPT
- Want to ringfence dataset for large companies
- Open source -> what thought would take until 2030 took a few weeks
- With ChatGPT can switch to other open source LLM when good enough
- Can attract funding for proof of concept projects
- Elicit.org -> tool which have indexed all public journals
- AI isn't going to take your job but the job of the person not using AI

- Businesses have programs which can transform data into formats which can be used by LLMs
- Healthcare workers collecting certain data for specific questions
- Need to have the factors to measure the data
- The data collected is from what the person is thinking not in terms of machine learning

**SO
DU
2023**

SODU 2023

Day 1 - Session 2 - Room C

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Data Governance

Leader(s)

- Name: Julian

Session notes

Data Governance - what we collect how we collect and the Ethics of that data collection within local government

Julian from Open Data Manchester - 2010 community from all areas that wanted to see Good done by data. 'How can we get local government to release data so that people can understand their data around them' Buildings, new stories etc. Over time have matured from Why to why some data shouldnt be open and the decisions around why data cannot be open.

Greater Mancher approx 2.9 million, ten local authorities, combined authority has strategic overview of what happens in the area. Part of the governance of the combined authority was an Industrial strategy and around data, so IG and data governance forms a big part of 'We want more data' i.e what mechanisms are being putting in place to do that.

Information Board: Made up of people from health, police, local authority , private companies. And is an advisory board that decides what should be released, how etc. It has a good overview of the supply

and demand of data. Therefore from there we can work out a strategy. Currently thinning around the AI sector.

How can this space be created in local authorities so that its not just local authorities that make decisions around Information Policies.

Question - Is there any automated decision making - no one will admit to it but are aware of one, but one of the issues is people purchase systems and don't necessary know whether there is an AI element. So part of the role of the **IB** is to advise on this to allow local authorities to understand. AI appears to be being used in Decision Making within Children social care. The technologies are not making the decision themselves, AI pulls out the case load data and presents to case worker and then the liability is on the case worker. And so concerns around the nuance of the data and an understanding of AI. Fairness and agency of AI. It provides a space.

How focused are they on data quality? Pretty similar where ever you go? Data can't be used, rubbish in rubbish out.

Might be something similar in Glasgow

No members of the public on the board.

Every one's problem no ones responsibility.

At ODI summit - data having different value depending on where you are. Eg Telecomms data, connectivity, 4G - open signal 54th in the world. Ofcom, have their APIs for connectivity data, but its not easy to access and its restricted.

The above comms example would not really be bought into Information Board unless it was around Digital Inclusion.

Procurement: There is a disconnect in thinking that data is thought of last, when really we are paying for a service and it should be clear at the start that data is needed etc.

Health just looks at Health

There is a shared infrastructure, within Health, and they are represented on the board, they work slightly different way in Information Governance, but the ethics around how data is managed,

The board also has a devolved budget looking at transparent data sharing across the boards. Had to become an entity and see value in it so that it could get a budget.

More broadly in Scottish Government - where is the money going to come from to solve the problems. There is an argument of who is going to pay for it when it all goes wrong…

Generally do a DPIA - what are the processes in place and what are the mitigations and works in the GPDR, need to be aware of the law - its a complex involvement, and there is article 22 if there is an impact on certain characteristics. What happens when it goes wrong? Who is then liable?

Basic Open data around grant publishing data should adhere to certain standards but there is still a huge variation across scales.

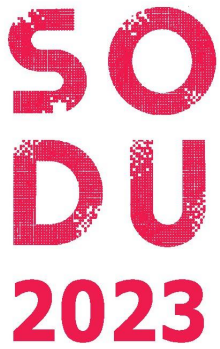Accessibility around data and there will be an investigation by Scottish Human rights.

IB has an opportunity to encourage community practice.

Fairness - within GPDR - the idea of what you loose by not doing something - risk of

Lots of AI examples and will find out what it hasn't got covered, AI - moved on to Data

Boards needs fair representation. Scotland is quite centralised do lack representation.

SODU 2023

Day 1 - Session 2 - Hot Desking

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Case studies for Open Data in Financial Services

Leader(s)

- Name: Mary Drabble

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

We discussed a partnership between Natwest (RBS), Renfrewshire local authority and Edinburgh University's Smart Data Foundry, where bank account data, matched to postcode, could be combined with a variety of local authority datasets to point citizens to relevant services. This was demoed at the Scotsman data Conference in Edinburgh in October 2023.

Links below:
https://www.nationalworldevents.com/sdc-2023/

https://www.scotsman.com/business/in-full-the-scotsman-data-conference-2023-4361160

https://smartdatafoundry.com/

https://www.ed.ac.uk/edinburgh-international-data-facility/services/bespoke-services/smart-data-foundry

We extended the discussion to how Private Sector does / doesn't share it's data and what the barriers to sharing data might be. We looked at other use cases for Finance data in the context of positive outcomes for the citizen / society.

In the context of data products, we discussed who should be accountable for data quality and pros / cons for quality assuring published datasets.

We touched on ethical issues in relation to adoption of AI and some of the topics in scope for the UK AI Summit, on November 1st & 2nd at Bletchley Park. Links Below:
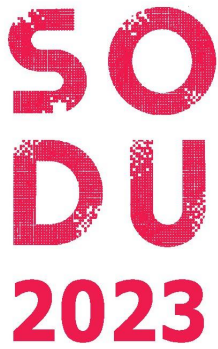https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about

https://www.aisafetysummit.gov.uk/

Contact Details for Mary Drabble: https://www.linkedin.com/in/mary-drabble-0041842/

- Context:
    - However, lots of funding for fintech projects
    - Generally for private gain — but real potential for public benefit!
- Case study #1: heatmap of spending behaviour by postcode (including breakdown by type of spending)
    - Use case: postcodes with sudden increase in overdrafts
        - These can be compared to service utilisation (e.g., foodbank use)
        - Offer tailored pop-ups (on bank app) or prompts (in bank branches) to direct customers to public services
- Case study #2: missing or forgotten pension pots
    - Allow beneficiaries to identify and reclaim their lost
- Case study #3: generating synthetic data using the publicly available data
- BCBS 239 rule mandates integrity of data used for making decisions for retail bank
    - Led to massive investment in data stewardship, data quality, data lineage
    - High maturity of data and pipelines — hence, greater confidence in making data available publicly
        - Due to burdensome regulations involving anonymisation, GDPR compliance
        - Also, due to (mistaken) belief in keeping a competitive edge by hiding data
    - What kinds of data is available
        - Pensions data, real estate sales, mortgages, default rate
    - However, no central repository
        - Bilateral contracts b/w smart data foundry and banks — hence no central standards
- New bill (2nd reading in Parliament)
    - Data protection bill

- ○ Not an open data bill, but seeks to expand/promote sharing data
- Change in industry attitudes:
  - ○ E.g., oil industry kept reams of data about oil well location, logging, etc. well-guarded and under locks. However, new regulations/legislation require public data sharing before applying for licence
- Concerns about commercial hanky-panky
  - ○ E.g., making only partial dataset available and charging for rest
  - ○ Ethical concern: selling of customers data and lack of benefits for that person
- Royal Bank of Scotland running a project to make data publicly available:
  - ○ To Smart Data Foundry
  - ○ Ultimately will be commercialised
- Lack of common standards in reporting of data to regulators — makes sharing of data b/w public and private actors
  - ○ E.g., by gambling shops
  - ○ Lack of standard taxonomy, indexing, metadata
- How to incentivise public bodies to release data openly
  - ○ And how to allow ppl to use the data
  - ○ Open nature of data means no tracking of use cases, stories told
    - ■ For instance, private companies don't openly declare their use of public data, which is commonplace
  - ○ Alternative is setting up a sign-up wall and demand potential use — however that risks compromising "openness" of data
- Who is responsible for sanitising/clearing data?
  - ○ Data producers vs consumers
    - ■ Or, intermediary creator of data products (i.e., who creates dashboards)
  - ○ Lack of teaching about ethics & governance to students in DS/AI education programs
  - ○ Fast-changing regulatory landscape
    - ■ Sticks (fines) are best for improving private players' behaviour including in sharing data
    - ■ However, data producers may prioritise complying with rules/laws over more practical issues re: data governance
  - ○ Having a Chief Data Officer may improve performance

SODU 2023

Day 1 - Session 2 - The Lab

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Local planning and environmental data - What's the Story in Torry?

- St Fittick's Park
- Just Transition Fund
- QGIS for mapping data

- [Datsette.io](#)
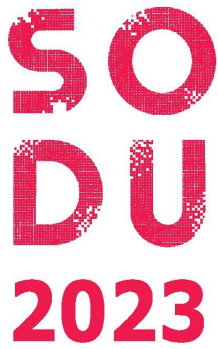-

Leader(s)

- Name: Jo Walsh

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Git: https://gitlab.com/ultrazool/story_in_torry
Local Campaign Website: https://saintfittickstorry.com/

- Looked at masterplan - traced areas into [QGIS](#) - digitised
- Planning data is not usually available as GIS data
- Requesting data and getting a licence from local authority takes time or isn't possible
- Tracing ends up being the easy route!...in 2023
- Datasette supports spatial/GIS (still in testing - bit buggy)
- Add Scottish Index of Multiple Deprivation? https://simd.scot
- Access planning permissions: [publicaccess.aberdeencity.gov.uk/online-applications](#)
- What are the positives? Jobs? Local Jobs?
- Local groups - share data/maps with local campaign group - get there thoughts

SODU 2023

Day 1 - Session 3 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Data Horror Stories

Leader(s)

- Name: Julian

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

How things can go wrong with data

Reinhalf and Rockoff(probably mispelt) Spreadsheet error problem

Paper released around 2005-2006 where analysed major world economies
Trying to find pattern between GDP and country's debt
2008 -> global collapse. Spread across the world and all economies effected
Referred to research
Used as an argument to put countries into austerity
Only 5 years ago degree student did economics degree and looked at research and original data in spreadsheet
One of the formulas was copied wrong meaning the analysis on the data was wrong
Excel spreadsheet error can lead to big consequences

Social care management system used to manage data

Email clients about change to service arrangements
Was going to go to everyone in database and would get stamped through mail machine (letters sent as well)
First letter on first pile contained sensitive information about the client
Spotted by mail room assistant who stopped all mail going out and erported it
Had to be investigated
Data manager used a text field in address line to store sensitive data about a person rather than using a flag in the database since the system wasn;t designed to hold that data

Company had to write to customers saying company name changed

Used surname field for person died
Sending letters to person of firstname and then deceased
Other ones on technical debt
Try to avoid this
3000 people with same birth date as it was used as a default
Address data very tricky and no real standard on how to format
Address form didn't collect country field to make easier for users (all in UK)
In data model, didn't default to UK field
Address dataset cannot be integrated anywhere as have no country field

Problems of AI going wrong

University developed project for what is edible or inedible fungi
Trained on edible mushrooms and inedible mushrooms
AI would try and work out what was specific about edible and inedible fungi
Tested in wild
AI started to give inedible outcome for edible mushroom
Found AI trained so it was edible mushroom and then inedible mushroom
Caused AI to think every mushroom was inedible

Large bank

Sent out to London
All that happened was work on Excel spreadsheet to see if data was right
Due to amount of money used was so big that couldn't have a mistake in the data
Oxbridge students looking at data

Looking at series of things picked for unique identifier

Used IBS101, IBS102, … IBS109
Allowed for 81 possible values
Will run out quickly
Data model items can get transferred into datasets
Be careful when thinking about putting numbers on something
Unique identifiers should have no meaning

Freedom of information requests

Large amount of personal data in requests in spreadsheet as problem
Complaints about antisocial behaviour
Wanted spreadsheet of data
Ended up sending data of what they thought was garbled nonsense
Turned out had things about email addresses and wanting to remain anonymous
Trying to figure out where chip shop called chippy and where chipper or chip shop
In feedback from people, had people selecting other because they wanted to say something slightly different from one of the options
Took a long time to fix

2016 Excel mangling data was standard identifier for genetics

Scientists had to change way data entered until Microsoft rolled out a fix

Train dataset to identify husky or wolf

Some of training data when showed husky was on snowy background -> something on snowy ground was a husky

People working on cleaning up metadata for archive who were old. Categorised all newspapers as English. Ended up ruining a report as language data was wrong

Problem with sex at birth and gender field. Datasets are merging the two and don't have the sex of birth. Data we want today may not be in the data source and could end up sourcing the wrong data.

NHS put out open data on website and starting training on them. During training, person who extracted the information hadn't done so correctly which could have ended up producing the wrong product without local knowledge. One person hadn't put good enough security on their database and the passwords could be accessed and read.

Manchester has tram system but fair structure was very opaque - hard to tell how much it costs to go between places

Extracted open data and got fair system
Got told from freedom of information request that the tram company couldn't release the information as their competitors might use it
Told copyright in dataset and couldn't release it
Release it with costs transformed into emojis
In the end transport system released their data

In Edinburgh, have new tram system with better colours on stops

Have timetable which shows the time it should be arriving but not when it is actually coming
Edinburgh can't use GPS system for boards for another 6 months even though it's used on the app

Transport data

Told 1st bus network would look at bus stop and see buses going to be here
Then they would vanish
Ended up ghost buses going around

Bus companies powerful
Don't release real time data but timetable
Done so the public can't complain about the times

Complaint from data managers can't release data as not accurate enough

Had data for schools catchment areas areas
Data not accurate enough as 5 houses missed from any area on map
Others were in two school's areas as shapes intersected
Instead of fixing data and releasing as open data, refused to release it

Library fill in survey every year

Looking at data
Question about how to support digital skills
Used some graphs
Appeared to be delivering training in area that didn't look right
Looked at question
Combined two datasets
Had one to one training in as well which wasn't offered
Need to think about collecting data that there isn't two datasets in same field

Word Cloud was around and being used for everything

Survey on staff in local authority - Wanted to see what staff saw as a problem
Resulting text put into word cloud
Managers all called into a meeting to get berated
Biggest problem was "management"
Analysis and data viz by word cloud had clumped "data Management", "project Management", "financial management", "time management" into one word.

Technology company

Lots of data rows
Had to release project recently
Wrong link made between dataset and pandas stashed it so took a long time to fix

TSB Data migration

went wrong and they were heavily fined by the Financial Conduct Authority (FCA) in Dec 2022. Read all about here:
https://www.fca.org.uk/news/press-releases/tsb-fined-48m-operational-resilience-failings

The 2 final notices issued by the FCA and PRA (Prudential Regulatory Authority) contain a very detailed description of the failures. Lots of lessons to be learned in there!

Data standards

Late 90s had Mars explorer
Sent on long journey
Going to orbit Mars and map possible landing sites
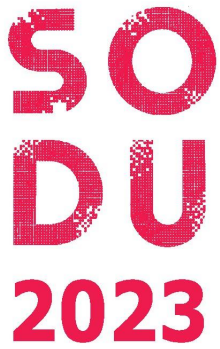Nose dived into Mars
Took a long time to find what was wrong
Mixed up imperial and metric units

Customer data

Single view of customer be afraid
Unique id in UK should be national insurance number but not always the case
Problems with merging data

SODU 2023

Day 1 - Session 3 - Room C

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session
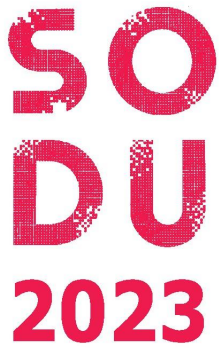
- Open Nap

Leader(s)

- Name: Jez and Katie

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Listened to https://www.youtube.com/watch?v=ymGrSgw0zHU

zzzzzzzzzzZZZZZZZZZZZZZZZzzzzzzzzzzzzzz

SODU 2023

Day 1 - Session 3 - Hot Desking

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- FAIR for Health Data

Leader(s)

- Name: Shahzad

Session notes

- Health Data Research (HDR) UK is an open data initiative

- Difficult to work across datasets:

  - Cannot query across databases (e.g., for males) — hence **findability** is a huge challenge

  - Requires approval from different boards

- Lack of common standards

  - In terms of file format, data structure, and semantics

- One possible solution: developing domain-specific common data models (CDMs) and transformers

- Another solution: developing "reproducible analytical pipelines" (RAPs)

- Semantics

  - Name of attribute: "health conditions" vs "issues" in two datasets

- ○ Within a field: 2 genders vs 5 gender categories
- ○ It affects **interoperability**
- CDMs
  - ○ OMOP — e.g., for pooling data from different regional health boards
  - ○ European-wide standards
    - ■ E.g., lack of a unique digital identifier for health records across jurisdictions
- "Open Data" in Health — is it open?
  - ○ Always requires institutional credentials and/or available to actual researchers only
  - ○ Yes, because "open" means as open as possible under law, and as a society we agree a different threshold for health records (which have a greater risk of abuse)
  - ○ No, because it's not open data. FAIR is a more suitable (and evasive description)

SODU 2023

Day 1 - Session 3 - The Lab

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Taxonomy for data

- Domain space - open data gov. https://find.data.gov.scot/
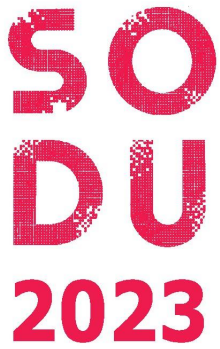- Publishes data using Schema json - to make it searchable in Search Engines (Google etc)

Leader(s)

- Name: Gautham

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- Deep web - data not discoverable?
- Taxonomy - wiki categories inspired categories/domains on find.data.gov.scot and europa
- Domains - high level categories
- Topics - suggested new topics
- Use cases / Keywords could be word cloud of datasets return by search query
- Feedback form for use case on site
- Add LLM - at datasets level initially

SODU 2023

Day 1 - Session 4 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Visibility into council structure and decision making

Leader(s)

- Name: Steve

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

App for where the seats come from

Next layer looking at the decision making hierarchy

Decision making committees and process and decisions made, where do they capture the logic around it

Political Leader

Executive Leader

There will be a report for a proposal for bus gates which would go to committee of the elective members

'We think….'

Voted on for by councillors

The elective councils will have a select meeting theme and can refer to full council

Is there a Hackathon that can build something

How can you get data around all these topics?

Democracy Club host info on elected councillors

HRI - platform - promote open data in Helsinki - do work around getting the data used. HRI developed [an App to track decision making](). Available as open Source. Could be replicated.

All local authorities are doing things, there is an Improvement Service in Scottish Government that was set up to improve local processes but not much done

Download all PDFs and feed it into GPT and then have a chatbox conversation

Aberdeen City Council publish a list of councillors who attended meetings. Not a record of who voted for or against a motion or proposal - just a count. Compare with Helsinki, above.

Previous suggestions should include who voted on it, they do count it but they do not track it.

Look at What do they know - for Aberdeen City Councils, put in hundreds of requests.

What are the local byelaws are in place - you have to FOI it. There are signs at the beach etc but have never been enforced.

Doing a local democracy hack will create pressure.

No one local authority present this weeken, no one from COSLA, SOLACE or most government agencies, There is an open government partnership - not here.

Still living under a legacy of doing services to citizens, and not for or better with them.

This should be designed for a 12 year old operating it.

Links

https://committees.aberdeencity.gov.uk/mgListDeclarationsOfInterest.aspx?XXR=0&DR=08%2f11%2f2022-19%2f11%2f2023&ACT=Find example of list of declarations

https://www.aberdeenshire.gov.uk/media/9499/GuidanceCouncillorRoles.pdf an example of committees in Aberdeenshire Council

https://committees.aberdeencity.gov.uk/mgMeetingAttendance.aspx?ID=8509 example of meeting attendance

https://committees.aberdeencity.gov.uk/mgCommitteeDetails.aspx?ID=334 shows a break down of an example of political parties in committee but not interactive

Aberdeen Councillors list: https://committees.aberdeencity.gov.uk/mgMemberIndex.aspx

Aberdeen City Council Structure:

https://www.aberdeencity.gov.uk/services/council-and-democracy/aberdeen-city-council-structure

https://www.aberdeencity.gov.uk/sites/default/files/2023-11/ACC%20structure%20chart%20Nov%202023.pdf

Council Committees:

https://www.aberdeencity.gov.uk/services/council-and-democracy/about-council/council-committees

Aberdeen City Health and Social Care Partnership:

https://www.aberdeencityhscp.scot/about-us/our-executive-team/

Outside Bodies: https://committees.aberdeencity.gov.uk/mgListOutsideBodies.aspx

Example Council Committee Attendance info:

https://committees.aberdeencity.gov.uk/mgAttendanceSummary.aspx

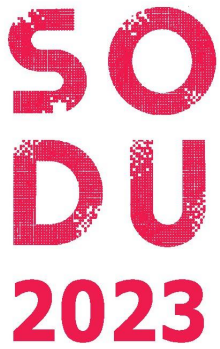Example Councillor attendance data: https://committees.aberdeencity.gov.uk/mgAttendance.aspx

Aberdeen Provost:

https://www.aberdeencity.gov.uk/services/council-and-democracy/civic-information/lord-provost-aberdeen

About the council: https://www.aberdeencity.gov.uk/services/council-and-democracy/about-council

Community Council Goverance information:

https://www.aberdeencity.gov.uk/services/council-and-democracy/community-councils/community-councils/community-council-governance

Council and Democracy information: https://www.aberdeencity.gov.uk/services/council-and-democracy

SODU 2023

Day 1 - Session 4 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Visibility into council structure and decision making

Leader(s)

- Name: Steve

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

App for where the seats come from

Next layer looking at the decision making hierarchy

Decision making committees and process and decisions made, where do they capture the logic around it

Political Leader

Executive Leader

There will be a report for a proposal for bus gates which would go to committee of the elective members

'We think….'

Voted on for by councillors

The elective councils will have a select meeting theme and can refer to full council

Is there a Hackathon that can build something

How can you get data around all these topics?

Democracy Club host info on elected councillors

HRI - platform - promote open data in Helsinki - do work around getting the data used. HRI developed [an App to track decision making](#). Available as open Source. Could be replicated.

All local authorities are doing things, there is an Improvement Service in Scottish Government that was set up to improve local processes but not much done

Download all PDFs and feed it into GPT and then have a chatbox conversation

Aberdeen City Council publish a list of councillors who attended meetings. Not a record of who voted for or against a motion or proposal - just a count. Compare with Helsinki, above.

Previous suggestions should include who voted on it, they do count it but they do not track it.

Look at What do they know - for Aberdeen City Councils, put in hundreds of requests.

What are the local byelaws are in place - you have to FOI it. There are signs at the beach etc but have never been enforced.

Doing a local democracy hack will create pressure.

No one local authority present this weeken, no one from COSLA, SOLACE or most government agencies, There is an open government partnership - not here.

Still living under a legacy of doing services to citizens, and not for or better with them.

This should be designed for a 12 year old operating it.

Links

https://committees.aberdeencity.gov.uk/mgListDeclarationsOfInterest.aspx?XXR=0&DR=08%2f11%2f2022-19%2f11%2f2023&ACT=Find example of list of declarations

https://www.aberdeenshire.gov.uk/media/9499/GuidanceCouncillorRoles.pdf an example of committees in Aberdeenshire Council

https://committees.aberdeencity.gov.uk/mgMeetingAttendance.aspx?ID=8509 example of meeting attendance

https://committees.aberdeencity.gov.uk/mgCommitteeDetails.aspx?ID=334 shows a break down of an example of political parties in committee but not interactive

Aberdeen Councillors list: https://committees.aberdeencity.gov.uk/mgMemberIndex.aspx

Aberdeen City Council Structure:

https://www.aberdeencity.gov.uk/services/council-and-democracy/aberdeen-city-council-structure

https://www.aberdeencity.gov.uk/sites/default/files/2023-11/ACC%20structure%20chart%20Nov%202023.pdf

Council Committees:

https://www.aberdeencity.gov.uk/services/council-and-democracy/about-council/council-committees

Aberdeen City Health and Social Care Partnership:

https://www.aberdeencityhscp.scot/about-us/our-executive-team/

Outside Bodies: https://committees.aberdeencity.gov.uk/mgListOutsideBodies.aspx

Example Council Committee Attendance info:

https://committees.aberdeencity.gov.uk/mgAttendanceSummary.aspx

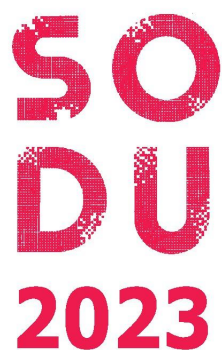Example Councillor attendance data: https://committees.aberdeencity.gov.uk/mgAttendance.aspx

Aberdeen Provost:

https://www.aberdeencity.gov.uk/services/council-and-democracy/civic-information/lord-provost-aberdeen

About the council: https://www.aberdeencity.gov.uk/services/council-and-democracy/about-council

Community Council Goverance information:

https://www.aberdeencity.gov.uk/services/council-and-democracy/community-councils/community-councils/community-council-governance

Council and Democracy information: https://www.aberdeencity.gov.uk/services/council-and-democracy

SODU 2023

Day 1 - Session 4 - Room C

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- How best to prepare yourself for an Open Data Job
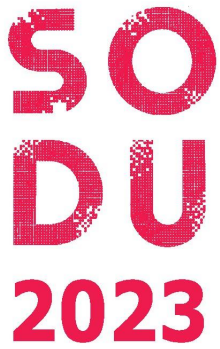
Leader(s)

- Name: ISI

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- Prepare for interviews — be authentic, creative

    - Interpersonal skills

    - Project management
      (e.g., own wedding planning as an example)

- Networking is key: "there's no substitute for face-to-face chats"

- Job requirements are often aspirational — don't be intimidated!

- Technical Career stages in Data:

    - Data analyst (no code, Excel)

    - Data scientist (knows ML)

    - Full-stack DS (autonomous)

- DAMA offers a certified data management professional qualification (CDMP) and mentorship opportunities. Student membership is available: https://www.dama-uk.org/Membership-benefits

- Data and Tech intersection has a ton of job titles (see DAMA blue book - DMBOK):

    - Data governance, integration, engineer, modelling, security, privacy, etc.

- Linked-in is your friend. Do create a profile even if you don't 'interact' with Social Media. It has a great job search functionality and nearly all professional employers advertise their vacancies on Linked-in.

- Some jobs may get over 200 applications (esp in Public Sector). Don't give up. Keep applying. Try and seek feedback where you can.

- Some companies specialise in Early Careers and aim to bridge the training gap between academic study and application to industry. This is one example: https://www.fdmgroup.com/careers/

- For early careers employers don't actually care that much about 'what you studied' - it's how you study that's important. What's your approach to problem solving? Do you have good interpersonal skills? Are you a quick learner? Are you open to new ideas? Etc.

- Many people who work in 'data' do not have 'data' in their job title. Data is everywhere and data management skills are essential to the operations of most companies.

SODU 2023

Day 1 - Session 4 - The Lab

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Crowdsourced data - How do we index it?

Leader(s)

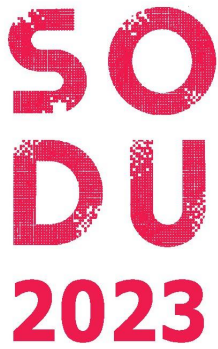- Name: Jack Gilmore

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- Feeds that conform to an agreed schema
- What is the difference between crowdsourced vs. officially published?
- Examples of crowdsourced data
  - Open Plaques
  - Open Benches
  - Aberdeen Harbour arrivals
- Crowdsourced datasets should be of good quality and correctable too
- Crowdsourced dataset scale
  - Birds **I've** seen in my garden
  - Birds everyone has seen in their gardens around the UK
- Suggest for people how to publish and conform to our metadata standards
- Motivations for publishing a feed
  - Prevent scraping
  - Dissemination and sharing of projects / data
- What is the purpose of Open Data Scotland?
  - "To make open data findable and accessible in Scotland"*
  - * Resources allowing
  - Therefore, we should have the crowdsourced datasets indexed on our website!
- Process for submitting datasets for indexing

- Agreed standards
- Minimum viable metadata

SODU 2023

Day 1 - Session 5 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

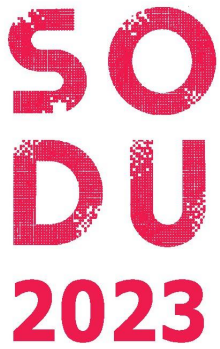## Title: GitHub Actions - What do they do?

Leader(s)

- Name: Jack Gilmore

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone
- Core use case for GitHub Actions
  - Automated task for publishing your website
- DC Thomson
  - A&E waiting times
  - https://github.com/DCTdatateam?tab=repositories
- Continuous Integration / Continuous Deployment
  - Running unit tests on code
  - Deploying apps
- Open Data Scotland pipeline
  - https://github.com/OpenDataScotland/opendata.scot_pipeline
- Covid publishing dashboard
  - https://github.com/LaurenceMolloy/DataEngineeringSandbox
  - https://laurencemolloy.github.io/DataEngineeringSandbox/docs/UKCovidDashboard
- Watch out for 60 days inactivity timeout
-

SODU 2023

Day 1 - Session 5 - Room C++

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Weather

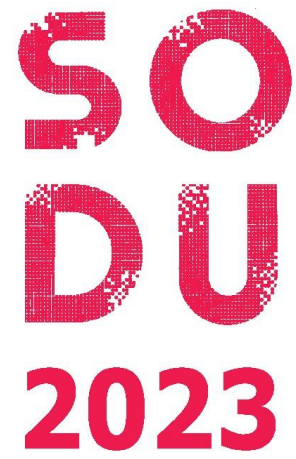## Leader(s)

- Name: Steve and Karen

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone
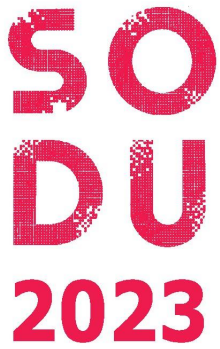
https://www.wunderground.com/wundermap?lat=57.204&lon=-2.194

https://energysavingtrust.org.uk/

https://www.wxcharts.com/?panel=default&model=gfs,gfs,gfs,gfs&region=sco&chart=overview,850temp,wind10mkph,snowdepth&run=12&step=150&plottype=10&lat=57.148&lon=-2.093&skewtstep=0

SODU 2023

Day Two

19 November 2023

SODU 2023

Day 1 - Session 5 - Hot Desking

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

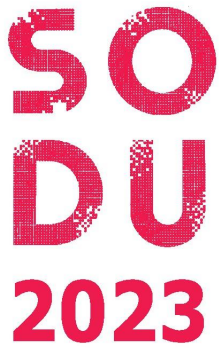- **Open Data Séance**

Leader(s)

- Name: Ian

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- How to contact those who aren't with us
- Was a team in every authority working with open data but not here
- Open Data Camp UK (in Aberdeen) 2017 were people from the local authority
- How to make this event attractive to people who aren't here?
  - Perhaps it is at the weekend
  - Seemed more like a convenient excuse
  - Might be an element
- Could be a funding issue
- May act as a barrier for some people
- Good if this is paid for by a company
- People may like the event to come to them
- Companies may hide the fact they use open data
- May be good to get in front of more different events
- Nobody from the local council
- Can be refreshing to not have the event from the same places
- Hierarchical culture in government
- Could be a chicken and egg problem
- With more people may have more people presenting

- Data is [open by default in Scotland](#)
- Have been talking to the Government about open data for a while now
- Have 3 people working on being open and transparent about the Government
- How do we contact them?
- Some people's paths haven't crossed before even though they have been to the same events
- How to promote event earlier in advance?
- Scottish Government doing open data all over the place
- There are lots of baskets over there (in terms of opportunities)
- It would be good to see what people are getting out of these events
- It would be good to make grant funding from the Government more effective
- It may be good to have a debate this issue kind of thing and to keep it up throughout the year
- Might be able to put the issues talked about at SODU online
- Might need a more targeted event
- Only 2 companies working fully in open data in Scotland. Neither are here
- Might be good to put open in front of things so people know its open data
- Open data as a category might be too broad
- Might make more sense to talk about data rather than just open data
    - Data unconference
- It can be hard for data startup companies South of Scotland to expand North - might be opportunities to be explored there
- Talk about what we do with the data here
- How to market and communicate better
- Also getting to the local community
- Make sure people know they don't have to be technically literate - can come and ask questions
- Councils see themselves as taking closed data and making it open while we are looking at what we do with that open data - may want to see how to change this narrative
- Couple of thousand people in councils who are driven by data
    - Would hope they would engage
    - May be good to have a lunch and learn
    - Some people have no context of the data they're using
    - Councils are overstretched
    - May see data through a lens of their job rather than it could be helpful for the general public
- Lots of sectors with their own data and wouldn't share it and it was making their life worst
- May need to make people understand that for open data the cost of inactiveness is worse than doing it
    - Is a shift in mindset
- May be good to look at the attendance of things the ODI sponsor
- Can ask people why they didn't come
- Open data was pushed by individuals who have mostly retired today
- Nobody has open data in their job description that we know of
- Evaluation is not on how much they've done in open data
- Now becoming something that is on the side
- People saying they don't have time for it and don't think they'll get anything from it
- Need to reward them for it
- Open data landscape massively undervalued
    - £2.21bn p/a should be generated
- Need funding for it and where is this going to come from?
- Why does the data lab not have a position on open data?
- Gaps in understanding of open data
- May be good to connect with people in the Scottish Government
- Good to have the broader public sector

SODU 2023

Day 2 - Session 1 - The Lab

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session
● ChatSODU - ChatGTP for Humans
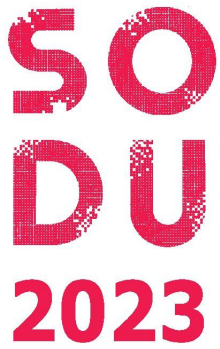
Leader(s)
● Name: Jez

Session notes
● Word game - each person says a word in a sentence - explanation of generating a sentence without having an understanding of the topic
● Fairytale word game example - expectation that this would start with Once upon a time.. (construct). Like family fortunes.
● ChatGTP we asked the whole english language internet. Predictive based on likelihood that a word would couple with another world.
● Search for all the documents that reference tapirs and fairytales – connected response.
● Doesn't always take the most likely word as that would be boring.
● Sometimes it will get stuck in a loop.
● Makes stuff up as it just says likely things. Mansplaining API :-) Bases hallucinations in plausibility
● Combine with a database search then you could be sure it is actually looking things
● Constantly moving in this area so this will change
● Articulate and confidence of the model leads to this being believed.
● Generative adversarial models are now being built into the model. This is where the danger is.
● Need to check the work.
● Tool that is a master at language
● Source identification and citing AI.
● Linking this to how we talk about usage of wikipedia.
● Need to be able to empower people to use these tools.

- Need to think about the data that you add to the model – implications for IP
- Thinking about how to assess knowledge and how knowledge is created. Example shared about how students are using ChatGTP to help with grammar and language constructs.
- Flooding the ecosystem with information
- Use the tools to help understand the validity of the information and check the data.
- Volume and creation of information and pollution of the knowledge ecosystem
- Real time translation and the empowering.
- When should you and shouldn't you use these models.
- Is there a discipline difference for when you should and shouldn't be using generative AI eg science vs humanities
- Using automation to enable creativity and criticality. Need to think about the use of gen AI in the context of the use case
- Need to be mindful of the economic and reform narrative to create efficiencies but need to be mindful of the bias and structural inequalities.
- Speed has increased very quick in a short space of time.
- Testbed and softwares never fully developed. Displacement of jobs and roles and activities.
- Forward looking view for how we use these technologies.
- Certain understanding and approaches. Asking questions.
- Links to pervious technologies and how this changed how things worked/how we approached things.
- Need to look at how this impacts and devalues work. This is impacting the professions.
- Paper by Xiang Hui, et al. The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4544582

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

SODU 2023

Day 2 - Session 1 - Room C

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Introduction to Wikidata

Leader(s)

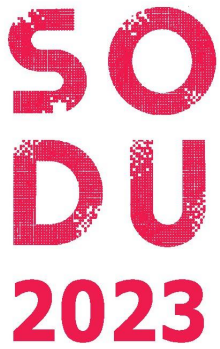- Name: Ian Watt

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Ian gave a talk he's given to University post-grads and other groups

Ian's slides:
https://www.dropbox.com/scl/fi/n2tg56afzkvsjdpwhbruw/IW_Wikidata_RGU.pptx?rlkey=thipmirdqzghb94gs0kpix7u4&dl=0

- May be good to look at it from a personal values point of view (personal motivation)
- How do decisions impact your daily life?
- Need data integrity
- Could be used in marketing
- Could be seen as part of your personal project
- Everything is theoretically possible
- Can be a part of personal marketing - trying to look into different data careers and whether there is an open data group in Scotland
- Over 25000 charities in Scotland
  - Could learn how to use data better
- May be a matter of time before people become interested
- Have the AI revolution which is making people more interested in data
- Perhaps make it about AI

SODU
DU
2023

SODU 2023

Day 2 - Session 1 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

● Running events for school pupils to learn about data

Leader(s)

● Name: Nevil

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Nevil previous teacher ran an event - got a lot of people from the data science community alongside 17-21 year olds. Went to every Statistical gathering in Edinburgh - do you want to be a mystery guests. Had 10 guests from statistics, criminology, health etc. Got them all to stand in a circle and had 40 students which broke in to 3-4 groups and spent 8 minutes with them. Had to figure out what they did they couldn't ask why they did, they had to ask questions. Hard to figure out the name of the job. If we did it in the next section - what do they do with job data (job titles were a bit confusing). Feed back at then end - mystery guests liked talking to students.

Maths week scotland - jobs of the unknown will show the information.

In person event - flows much much better - had lots of chocolate.

Diversity - forum - valuing data differently where you - i.e tele comms

Teachers coming long

Use cases and story telling level is probably some of the useful information, its maybe not just the coding, its understanding the applications.

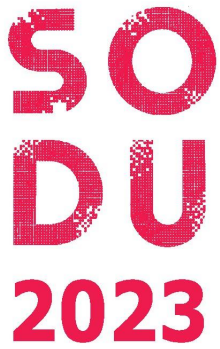Re-use resources from in-person events to make a toolkit?

Gamification / speed dating approach would really engage kids.

Recruitment campaign - don't be put off by the word of Statistician - all examples, homelessness, environment, health, communications,

A good way to get people thinking about data and AI is to help them realise it crosses over with lots of different areas of life. People are also good at spotting patterns, and that's what algorithms do too.

Folder of resourses used for the Sep 2023 event:
https://drive.google.com/drive/folders/1aJBrRnTiYM-6CoROZHSvFt6q8E99VzoK

SODU 2023

Day 2 - Session 2 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

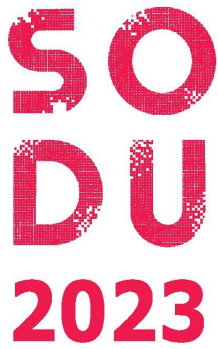- What is data journalism

Leader(s)

- Name: Lesley-Anne

Session notes

- Data journalism is now becoming part of the training for journalism students. Move from the previous attitude of becoming a journalist as for writing not numbers.
- General training also taking place within the journalism profession to develop data literacy skills.
- Development and creation of 'tracker articles' forms a core part of the role. 'Tracker articles' often have a high volume of traffic eg Covid tracker most viewed article across news providers.
- DDJ requires journalism, data science and design tools. Part of this is understanding design and visualisation eg emotions evoked by colours
- Examples - waiting times, data from PHS, storm Babet, data from SEPA, vacancy rates in Dundee shopping centres with data gathered by the team manually
- Not all automated, also a need to hand draw and gather the data, this is an approach that needs to be taken when data isn't granular enough or is just not available.
- Baby names data and articles:
    - https://www.thecourier.co.uk/fp/courier-investigations/3156332/scottish-baby-names-going-extinct/
    - https://www.thecourier.co.uk/fp/courier-investigations/3135287/baby-names-scotland/
    - https://www.thecourier.co.uk/fp/courier-investigations/3168365/baby-name-rules-scotland/
- Limited number of data journalists in Scotland but across the industry there is an increase in DDJ roles. Have seen an increase in Scotland but it is challenging to get resources in this area.

- Open data can help in terms of access and ability to report on a range of topics. Some data might not be beneficial for the creator but is beneficial for democracy and enabling transparency eg NHS Scotland wait time data.
- Freelance data journalism is an option for anyone interested in getting into DDJ. Useful vehicle to raise profile could be a blog.
- DDJ is a combination of working with the data and speaking to the experts to aid in the interpretation and to contextualise the data. The need to engage with experts is particularly important to engage with experts to understand the data.
- How can Open Data Scotland improve services? Need for the identification of superusers and release notes for changes to data.
- All data used is currently open source as are the data analysis tools. Concerns over licencing of commercial data and reuse.
- Caution on the use of small data sets and interpretation of these.

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

https://drive.google.com/drive/folders/15aaXAx8Nkv2jouyZuau4oDEGOd6ZPl1M?usp=sharing

SODU 2023

Day 2 - Session 2 - The Lab

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Using LLMs to compose and automate FOIs

Leader(s)

- Name: Jack Gilmore

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- One of the restrictions is around the number of hours (18-24 staff hours)
- Eg. Scottish Government - can't run expenses for less than £500 for a year but could do monthly so do we do monthly requests or ask for them to be put on Open Data.
- https://www.whatdotheyknow.com/ appears in public - even if they just send it back to you. Disclosure logs are not a requirement for all to maintain/publish. Using this site means that it becomes public record. In order to stop abuse though, you can only have a limited number of requests, you can get a pro account that allows you to do multiple requests.
- Question - if you use whatdotheyknow do you have to have a name, there is a requirement for there to be the name, but you don't need to provide a full name.
- Review process for FOI - should be answered within 20 minutes.
- Some people eg journalists don't put through whatdotheyknow as they will potentially want to scoop it.
- There is a need for automation and different names but different text, but want to do it randomly.

What if people present submitted a single request each for a month?

Volunteer	Date range
Jack	October 2023 https://www.whatdotheyknow.com/request/electronic_purchasing_card_data
Ian	September 2023 https://www.whatdotheyknow.com/request/epc_expenditure_below_ps500_in_s
V Snake	August 2023 https://www.whatdotheyknow.com/request/credit_card_purchases_august_202
John	July 2023 https://www.whatdotheyknow.com/request/small_bank_card_spends_in_july_2
K Stott	June 2023 https://www.whatdotheyknow.com/request/credit_card_spending_under_ps500#outgoing-1553577
Adrian	May 2023 https://www.whatdotheyknow.com/request/request_for_information_on_epc_e
Oliver	April 2023 https://www.whatdotheyknow.com/request/spending_via_epc_for_april_2023
LLH	March 2023 https://www.whatdotheyknow.com/request/request_for_information_re_epc_e

https://www.whatdotheyknow.com/request/epc_expenditure_under_ps500#incoming-2411331

https://www.whatdotheyknow.com/
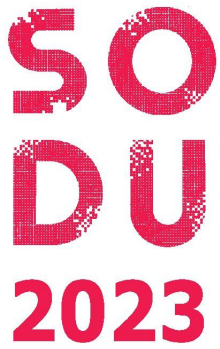
Dear Scottish Government,

Please provide the following information:

A report of ePC spends under £500 for **MONTH_GOES_HERE**, including:
- Directorate who performed the spend
- Merchant name
- Merchant category name
- Transaction date
- Transaction amount
- Expense description

If possible, please provide the data in a tabular format, preferably as CSV or Excel spreadsheets.
Please avoid providing the data as a PDF if possible.

Yours faithfully,

SODU 2023

Day 2 - Lunch - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Scottish Government - Open Gov Action Plan

Leader(s)

- Name: Louise

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Find out more in the blog - [Foundations first – realising the potential of open data - Digital (blogs.gov.scot)](#)

- Scottish Government Data Division - advice, policies and initiatives that help the Scottish public sector be more data driven and collaborative.
- Wider Data Domain includes official statistics and digital citizen team.
- Technical data policy team - open data & open gov, data innovation, data maturity, data standards, community of practice and leadership
- Open Government Action plan - Open Gov is about making governments more transparent, accessible, inclusive and participative. So governments don't just provide info, but help people understand info in a meaningful way.
- Involving people - improves decisions, which helps develop better policies. Services are improved because people have been involved. People feel involved/listened to, which helps generate trust between a gov and its citizens.

  Commitments are
  - Fiscal transparency
  - Health and social care

- Climate
- Participation
- Digital and data

- Within the digital and data commitment there are six activities
    - Data discovery find.data.gov.scot
    - Data maturity and developing a data transformation framework for data improvement in public bodies
    - Improving statistics.gov.scot
    - Community of practice for open data and data standards - includes recent work with recommendations for taking forward open data in Scotland (report will be published)
    - Opening up data relevant to other open government themes
    - Scottish AI Register- Scottish AI Register

Discussed

Can we make orgs provide more data?

How long will this take - do we have good examples of where open data has been used - lots of the examples are abroad.

Need to get out and talk the communities using data

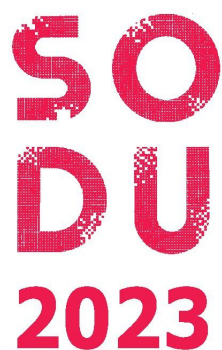Data maturity - this approach doesn't always resonate with people.

Links

Digital Foundations first – realising the potential of open data - Digital (blogs.gov.scot)

Digital Update from the pilot Public Engagement Panel - exploring data ethics for public sector data. - Digital (blogs.gov.scot)

Open Government - gov.scot (www.gov.scot)

Contact: louise.meikleham@gov.scot

SODU 2023

Day 2 - Lunch - Room C

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- Imaging a data discovery engine
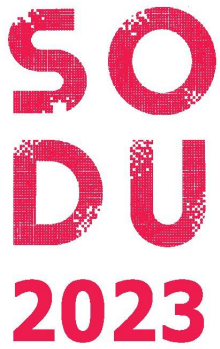
Leader(s)

- Name: Gautham

Session notes

- Change in user expectations for search now we have genAI
- Data for training LLM may not be inclusive of datasets discoverable Dtective
- Issues with data being behind a log in reducing the ability to preview data on Dtective
- Prompts to improve search*
- ChatGTP summaries of responses.
- Use of ChatGTP to understand how to construct a search eg drug abuse example
- Advanced search queries and use of boolean searching would be useful*

*most desired ideas

SODU 2023

Day 2 - Lunch - Hot Desking

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

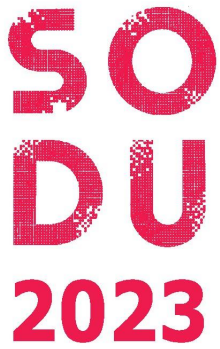## Title of session

- Github Actions - What do they do?

Leader(s)

- Name: Jack Gilmore

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

- Great way of having a website or page and having actions update it in some way
- Have actions triggered
    - Manually
    - Scheduled (cron)
    - Triggered by some action
- Can be used for pretty much anything
- You get free allowance, but some things are prohibited (eg bitcoin mining)
- Free for OS projects with generous allowances
- Great for hosting things such as https://opendata.scot
- Like containerised instances- import Python etc as you need it
- Github has demos and sample workflows
- Lots of examples and tutorials
- Steal from others - make changes - see if it breaks
- Used in CI / CD pipelines
- Companies and others have published actions that you can import and use
- You can create and host your own actions to import too

SODU 2023

Day 2 - Session 3 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

OpenStreetMap (OSM)

- Enhancing your data with OpenStreetMap
- Identifying Low Carbon Energy Opportunities

Leader(s)

- Name: Tim & Steve

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

[Continuum Industries](#) Business using open data to identify development opportunities

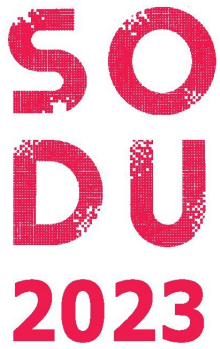https://fozy81.github.io/opendatascot/articles/riding-a-bike.html - Glasgow Cycle Trip data augmented with OpenStreetMap routes

Amazon contributing to OpenStreetMap for mapping edits are funneled through a single user accountant - https://www.openstreetmap.org/user/afurquan

Edinburgh History Mapping project info: https://www.ed.ac.uk/history-classics-archaeology/history/research/research-projects/mapping-edinburgh-s-social-history

History Edinburgh mapping tools: https://mappingedinburgh.org/tools/

SODU 2023

Day 2 - Session 3 - Hot Desking

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

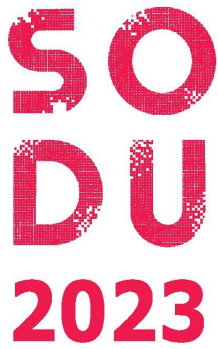- Lego for Reproducibility

Leader(s)

- Name: Varina

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Sorry we forgot to have anyone takes notes, but it was more of an experience anyway.
- Here's the original documentation for this workshop, under CC-BY license so you can use and modify for your own needs! https://eprints.gla.ac.uk/196477/

SODU 2023

Day 2 - Session 4 - Studio One

You can view details of all the sessions at https://bit.ly/SODU2023-SessionGrid

Title of session

- LLMs and Open Data - practical session

Leader(s)

- Lawrence Molloy (online)
- Steve Shearman (in person)

Session notes

All sessions are carried out under Chatham House Rules - please get permission before attributing someone

Watch the video of this session

We have source data (May 2023):
https://publichealthscotland.scot/publications/monthly-prescribing-activity-data/

It has a very useful glossary document:
https://publichealthscotland.scot/media/23197/open_data_glossary_of_terms.pdf

There is also a page with historic data files by month:
https://www.opendata.nhs.scot/dataset/prescriptions-in-the-community

Proposed goal: use chatGPT Playground to build a subject matter expert that can analyse the data to answer a simple data analytical question and provide the answer in graphical form (e.g. bar chart).
Question: How can we be sure that the answer we are given is accurate?