Ethical Considerations for Data Access and Use

Amy O’Hara, Georgetown University

Federal Committee on Statistical Methodology

Fall 2020

Overview

- What data are in scope?
- What data uses are in scope?
- Who is the public?
- What is engagement?
- Is engagement consent?
- Who is doing this?
- Goals and next actions

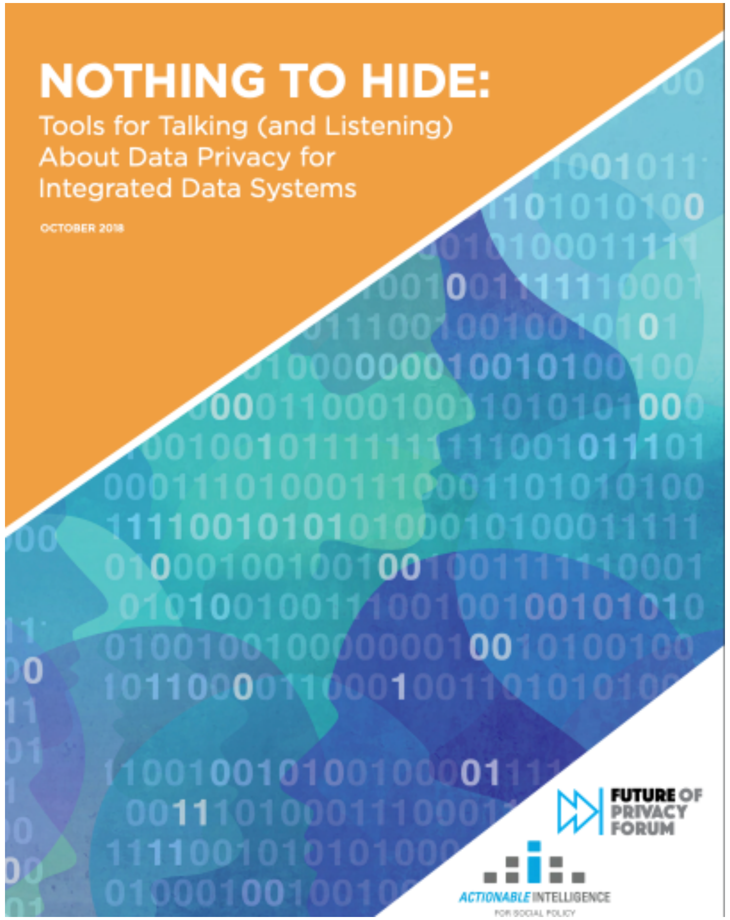
What data are in scope?

- Administrative data
- Not collected for statistical or research purposes
- Government and private sector

What data uses are in scope?

- Research, evaluation, evidence-building
- Beyond the administrative purpose for which the data were captured
- Government and private sector

Who is the public?



NOTHING TO HIDE:

Tools for Talking (and Listening) About Data Privacy for Integrated Data Systems

OCTOBER 2018

FUTURE OF PRIVACY FORUM

ACTIONABLE INTELLIGENCE FOR SOCIAL POLICY

C.7: Checklist — Public Engagement Meeting Planner

Having well thought-out meetings requires a lot of preparation. The following checklist will help you work through who, how, and what your meeting will address.

TARGET ATTENDEES / STAKEHOLDERS

☐ Community members
☐ Community leaders
☐ Government leaders
☐ Civic groups
☐ Other ______________

GETTING THE WORD OUT

☐ Social media
☐ Email
☐ Website
☐ Notification tree (using contacts to spread the word)
☐ Electronic invitation system (allows for collection of RSVPs)
☐ Existing publications (e.g., newsletters or newspapers)
☐ Posters or flyers

INVITATION MESSAGING AND FORM

☐ Clear, concise language written for the potential audience rather than bureaucrats or lawyers.
☐ Meeting expectations – What will the audience participate in and what can they learn?
☐ Contemplate and address in advance the potential attendee's question of why they should attend.
☐ Emphasize value of community member input.
☐ Double check the date, time, and location information.
☐ Consider whether supplementary materials will be distributed with the invitation.

WHEN

☐ Weeknight vs. weekend
☐ Morning, afternoon, or evening
☐ Do you want to send a message of inclusiveness? If yes, carefully consider your timing.
☐ Is there a holiday (religious or secular) that may conflict?
☐ Can this meeting be hosted multiple times?

FORMAT CHOICE

☐ Lecture
☐ Interview
☐ Mainly Q & A with short introduction
☐ Workshop or other interactive format
☐ Public hearing or debate
☐ Open house
☐ Facilitated learning
☐ Small group discussion or focus group
☐ Small group advisory board or expert panel
☐ Other ______________________________

Nothing to Hide: Appendix C: Engagement Worksheets, Checklists and Sample Materials | 39

Stakeholders

1. Government
2. Civic groups
3. Community leaders
4. Researchers
5. Data Intermediaries
6. Data subjects

https://fpf.org/wp-content/uploads/2018/09/FPF-AISP_Nothing-to-Hide.pdf

What is engagement?

Federal Statistical System	**AISP Integrated Data Systems**	**ADRUK Strategic Hub**
System of Records Notices (SORNs) Privacy Impact Assessments (PIAs) Federal Register Notices (FRNs) Advisory Committees Websites (online engagement)	Engagement Channels Social Media Website Publications (Newspapers, Magazines, Journals, etc.) Conferences Events Initiatives Universities and academic institutions	Initial engagement with voluntary, community and social enterprises, NGOs, etc. to raise awareness of projects Workshops or roundtables to gather detailed input on the proposed research questions and design, and flush out any overlooked issues Direct engagement with the public to seek input from people whose data will be used and whose lives may be affected, to hear about the issues important to them, and how they feel their data could best be used

https://fpf.org/wp-content/uploads/2018/09/FPF-AISP_Nothing-to-Hide.pdf

https://www.adruk.org/fileadmin/uploads/adruk/Trust_Security_and_Public_Interest-_Striking_the_Balance-_ADR_UK_2020.pdf

Is engagement consent?

JANUARY 14, 2020

Redesigning consent: big data, bigger risks

Over the last decade, the rapid proliferation of social media platforms coupled with the advancement of computational methods for collecting, processing, and analyzing big datasets created new opportunities for social science. But alongside new insights about the behaviors of individuals and groups, these practices raise new questions regarding what constitutes ethical research. Most critically, current disinformation scholars face a replication crisis exacerbated by uneven access to datasets, where social media users are unaware of their participation in academic research. Establishing scientific norms, where research is shared with the individuals whose data are accessed and processed in the name of science, involves redesigning consent and providing universal public access to databases. Ultimately, without methodological norms for disinformation studies, the field will continue to be dominated by corporate interests, further endangering the public's trust in disinformation research.

BY JOAN DONOVAN
Shorenstein Center on Media, Politics and Public Policy at Harvard University's John F. Kennedy School of Government

https://misinforeview.hks.harvard.edu/article/big-data-bigger-risks/

Companies, not people, should bear the burden of protecting data

David Medine and Gayatri Murthy · Wednesday, December 18, 2019

Privacy isn't dead as some would suggest ... but consent is. When was the last time you read a privacy policy for the apps on your mobile phone? Did you know that apps have privacy policies? How about reading the cookie notice on web pages you visit? Or reading the privacy notice on Internet-of-Things devices like your baby monitor?

brookings.edu/blog/techtank/2019/12/18/companies-not-people-should-bear-the-burden-of-protecting-data/

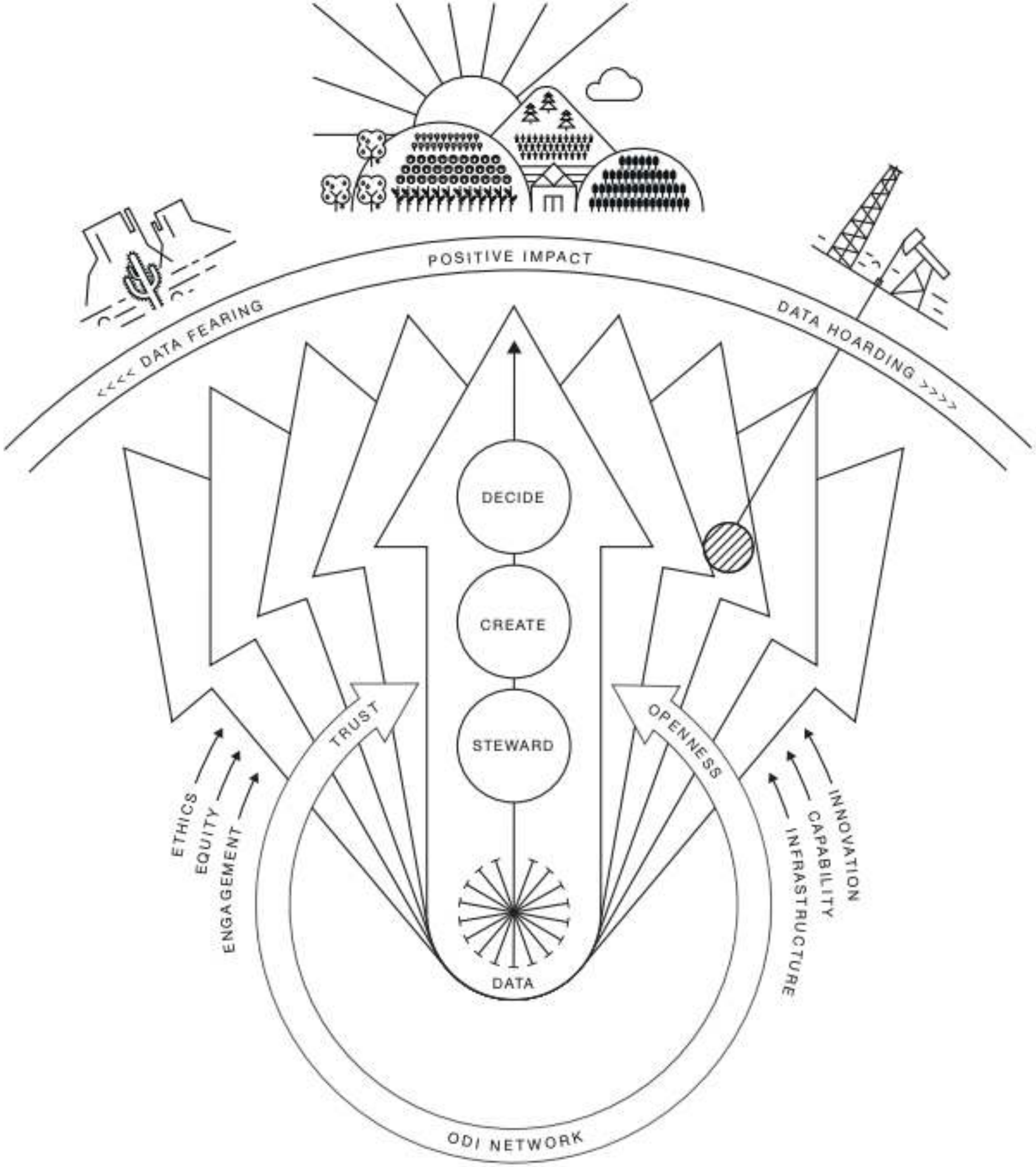
Tensions

Can we (legal) . . . . Should we (ethical)

Engagement . . . . Involvement

Notice . . . .   Consent

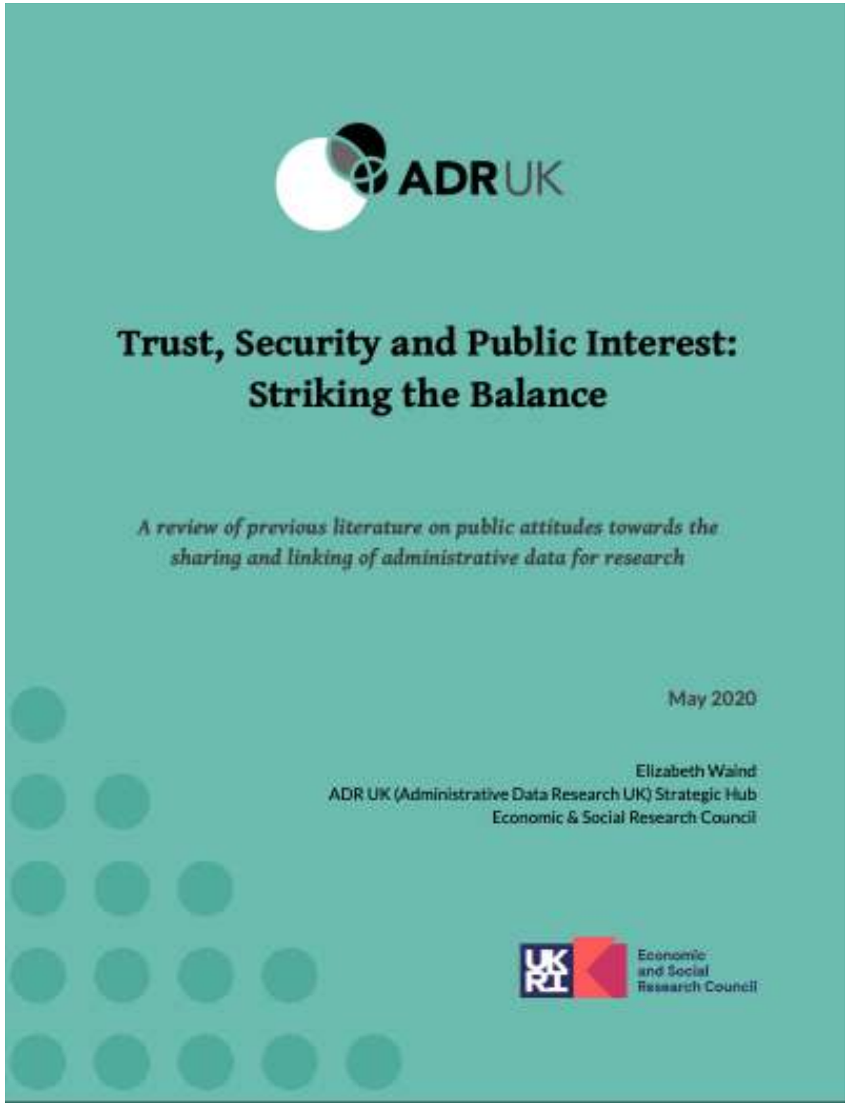
Awareness . . . . Acceptance theodi.org/theory-of-change



Examples? Who is doing this?

- Seattle Inclusive Outreach and Public Engagement Guide (link)
- City of Fort Saskatchewan (link)
- Digital Government New Zealand (link)
- Information Commissioner’s Office UK (link)
- Administrative Data Research United Kingdom (link)

Need this in US!



ADR UK

Trust, Security and Public Interest:
Striking the Balance

A review of previous literature on public attitudes towards the sharing and linking of administrative data for research

May 2020

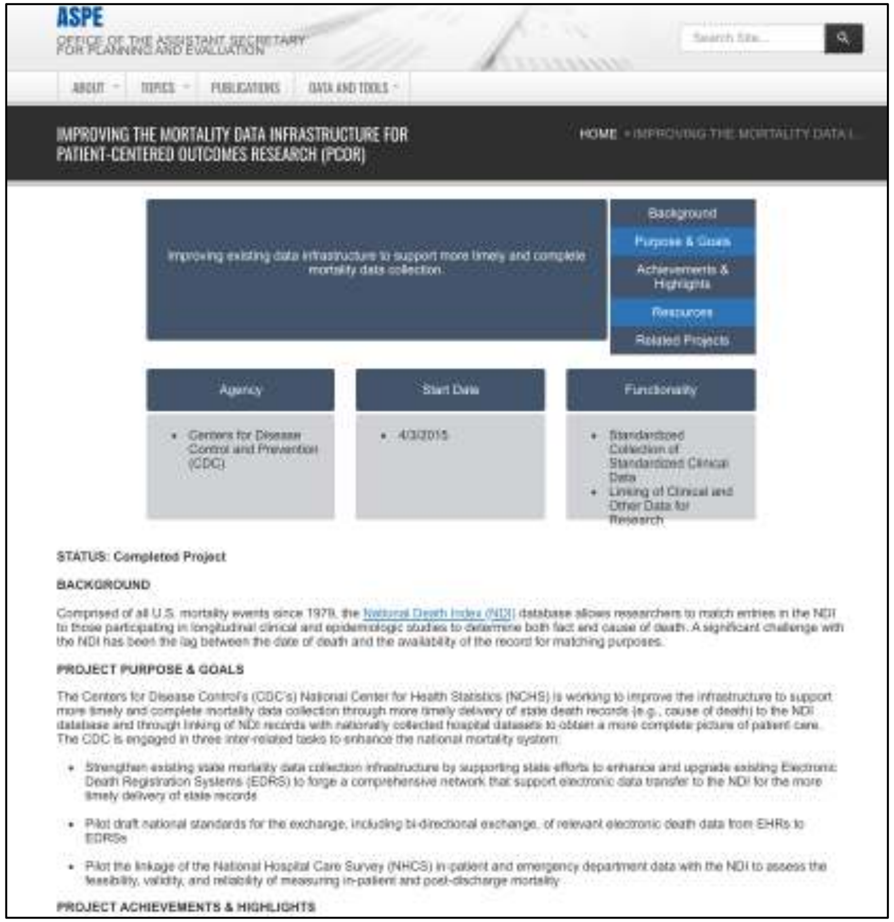
Elizabeth Waind
ADR UK (Administrative Data Research UK) Strategic Hub
Economic & Social Research Council

UKRI Economic and Social Research Council

Finds that the public is broadly supportive of their administrative data being used for research, as long as 3 main criteria are met:
1- Public interest: any research using administrative data is for the public interest and public benefit
2- Privacy and security: data being used must be de-identified, and protections put in place to prevent re-identification or misuse
3- Trust and transparency: trust of those holding the data is critical, and transparency around how that happens is critical.

https://www.adruk.org/fileadmin/uploads/adruk/Trust_Security_and_Public_Interest-_Striking_the_Balance-_ADR_UK_2020.pdf

US is already doing some things…



IMPROVING THE MORTALITY DATA INFRASTRUCTURE FOR PATIENT-CENTERED OUTCOMES RESEARCH (PCOR)

Improving existing data infrastructure to support more timely and complete mortality data collection.

Agency	Start Date	Functionality
Centers for Disease Control and Prevention (CDC)	4/3/2015	Standardized Collection of Standardized Clinical Data Linking of Clinical and Other Data for Research

STATUS: Completed Project

BACKGROUND

Comprised of all U.S. mortality events since 1979, the National Death Index (NDI) database allows researchers to match entries in the NDI to those participating in longitudinal clinical and epidemiologic studies to determine both fact and cause of death. A significant challenge with the NDI has been the lag between the date of death and the availability of the record for matching purposes.

PROJECT PURPOSE & GOALS

The Centers for Disease Control's (CDC's) National Center for Health Statistics (NCHS) is working to improve the infrastructure to support more timely and complete mortality data collection through more timely delivery of state death records (e.g., cause of death) to the NDI database and through linking of NDI records with nationally collected hospital datasets to obtain a more complete picture of patient care. The CDC is engaged in three inter-related tasks to enhance the national mortality system:

- Strengthen existing state mortality data collection infrastructure by supporting state efforts to enhance and upgrade existing Electronic Death Registration Systems (EDRS) to forge a comprehensive network that support electronic data transfer to the NDI for the more timely delivery of state records
- Pilot draft national standards for the exchange, including bi-directional exchange, of relevant electronic death data from EHRs to EDRSs
- Pilot the linkage of the National Hospital Care Survey (NHCS) in-patient and emergency department data with the NDI to assess the feasibility, validity, and reliability of measuring in-patient and post-discharge mortality

PROJECT ACHIEVEMENTS & HIGHLIGHTS

https://aspe.hhs.gov/improving-mortality-data-infrastructure-patient-centered-outcomes-research-pcor

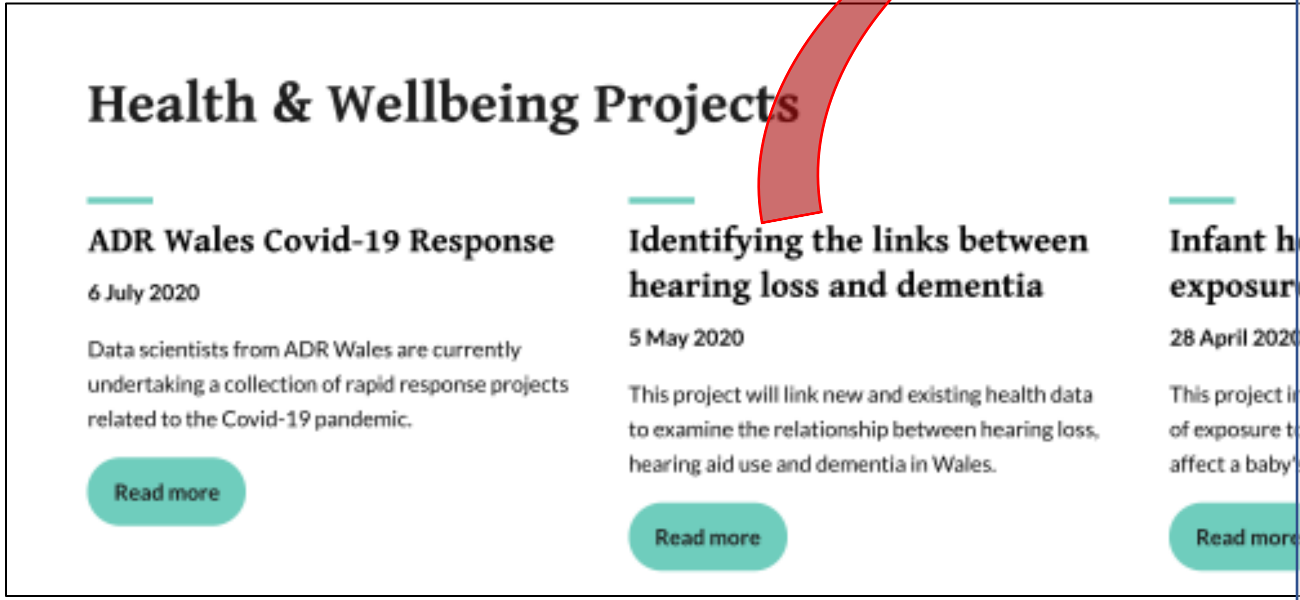
SOI Tax Stats - 2018 JSRP Selected Research Projects

2018 Statistics of Income Joint Statistical Research Program Projects

- Adding to the Databank: Creating Measures of Employer and Job Quality
- The Analysis of Benefit Program Interactions and Potential Outreach to Improve EITC Take-Up
- Analysis to Understand and Improve Taxpayer Compliance and Enforcement Related to Refundable Tax Credits
- Analyzing the Growth of Small Businesses, 2000–2016
- Behavioral Responses and Tax Compliance with Retirement Savings Contributions Limits
- Big Data Analytics and IRS Criminal Investigation
- Classification of Firms, Workers, and Return to Successful Innovation
- Creation of a Data Set of Unrealized Capital Gains and Haig-Simons Income
- Determinants of Corporate Organizational Structure

https://www.irs.gov/statistics/soi-tax-stats-2018-jsrp-selected-research-projects

We should aspire to this



Health & Wellbeing Projects

ADR Wales Covid-19 Response

6 July 2020

Data scientists from ADR Wales are currently undertaking a collection of rapid response projects related to the Covid-19 pandemic.

Identifying the links between hearing loss and dementia

5 May 2020

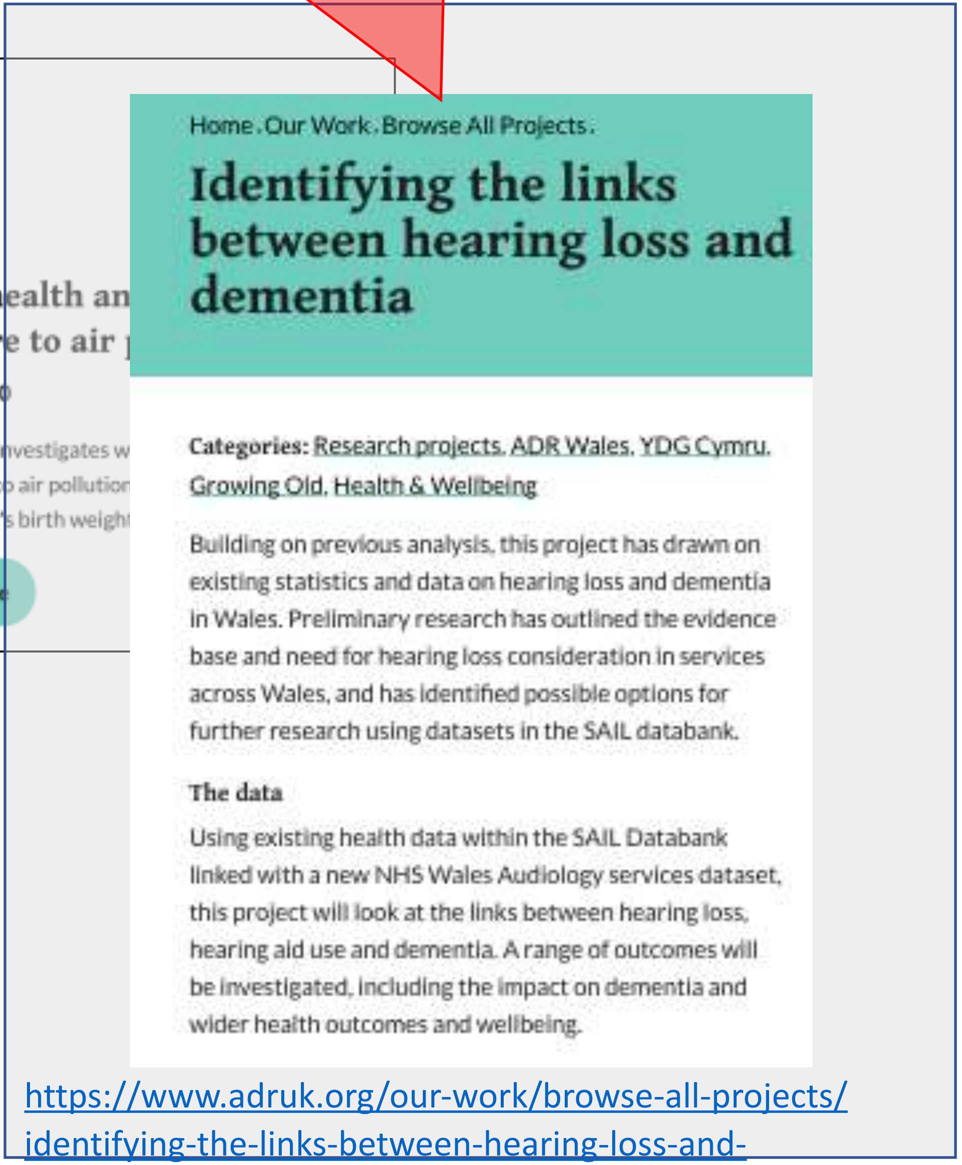
This project will link new and existing health data to examine the relationship between hearing loss, hearing aid use and dementia in Wales.

Infant health an exposure to air

28 April 2020

This project investigates w of exposure to air pollution affect a baby's birth weight

https://www.adruk.org/our-work/health-wellbeing/



Identifying the links between hearing loss and dementia

Categories: Research projects, ADR Wales, YDG Cymru, Growing Old, Health & Wellbeing

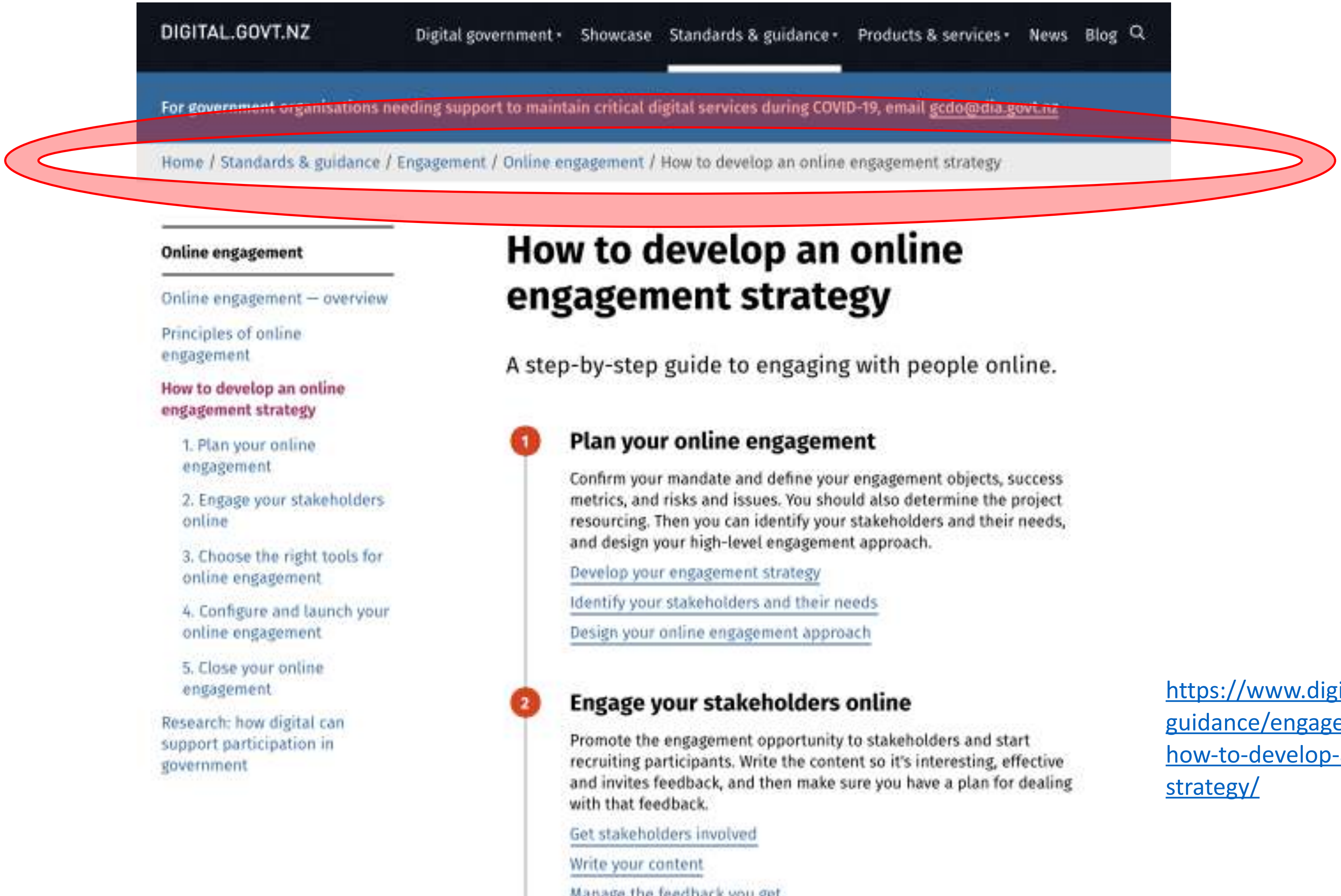
Building on previous analysis, this project has drawn on existing statistics and data on hearing loss and dementia in Wales. Preliminary research has outlined the evidence base and need for hearing loss consideration in services across Wales, and has identified possible options for further research using datasets in the SAIL databank.

The data

Using existing health data within the SAIL Databank linked with a new NHS Wales Audiology services dataset, this project will look at the links between hearing loss, hearing aid use and dementia. A range of outcomes will be investigated, including the impact on dementia and wider health outcomes and wellbeing.

https://www.adruk.org/our-work/browse-all-projects/identifying-the-links-between-hearing-loss-and-dementia-212/

Inspiration for Federal Data Strategy?



For government organisations needing support to maintain critical digital services during COVID-19, email gcdo@dia.govt.nz

Home / Standards & guidance / Engagement / Online engagement / How to develop an online engagement strategy

Online engagement

- Online engagement — overview
- Principles of online engagement
- **How to develop an online engagement strategy**
  - 1. Plan your online engagement
  - 2. Engage your stakeholders online
  - 3. Choose the right tools for online engagement
  - 4. Configure and launch your online engagement
  - 5. Close your online engagement
- Research: how digital can support participation in government

How to develop an online engagement strategy

A step-by-step guide to engaging with people online.

1 Plan your online engagement

Confirm your mandate and define your engagement objects, success metrics, and risks and issues. You should also determine the project resourcing. Then you can identify your stakeholders and their needs, and design your high-level engagement approach.

Develop your engagement strategy

Identify your stakeholders and their needs

Design your online engagement approach

2 Engage your stakeholders online

Promote the engagement opportunity to stakeholders and start recruiting participants. Write the content so it's interesting, effective and invites feedback, and then make sure you have a plan for dealing with that feedback.

Get stakeholders involved

Write your content

Manage the feedback you get

https://www.digital.govt.nz/standards-and-guidance/engagement/online-engagement/how-to-develop-an-online-engagement-strategy/

More inspiration for the FDS

Data sharing code of practice

Data sharing checklists

These two checklists provide a handy step by step guide through the process of deciding whether to share personal data. One is for systematic data sharing, the other is for one off requests.

The checklists are designed to be used alongside the full code and highlight the relevant considerations to ensure that the sharing complies with the law and meets individuals' expectations.

Data sharing checklist – systematic data sharing

Scenario: You want to enter into an agreement to share personal data on an ongoing basis

Is the sharing justified?
Key points to consider:

- What is the sharing meant to achieve?
- Have you assessed the potential benefits and risks to individuals and/or society of sharing or not sharing?
- Is the sharing proportionate to the issue you are addressing?
- Could the objective be achieved without sharing personal data?

Do you have the power to share?
Key points to consider:

- The type of organisation you work for.
- Any relevant functions or powers of your organisation.
- The nature of the information you have been asked to share (for example was it given in confidence?).
- Any legal obligation to share information (for example a statutory requirement or a court order).

If you decide to share
It is good practice to have a data sharing agreement in place. As well as considering the key points above, your data sharing agreement should cover the following issues:

- What information needs to be shared.
- The organisations that will be involved.
- What you need to tell people about the data sharing and how you will communicate that information.
- Measures to ensure adequate security is in place to protect the data.
- What arrangements need to be in place to provide individuals with access to their personal data if they request it.
- Agreed common retention periods for the data.
- Processes to ensure secure deletion takes place.

ico.
Information Commissioner's Office

https://ico.org.uk/media/for-organisations/documents/1067/data_sharing_checklists.pdf

Whose job is it?

- Research using linked, anonymized administrative data is already occurring extensively. Should public engagement simply be a matter of informing the public that this is taking place, or must there be a level of permission granted from the public to the researcher to use their data, even if it anonymized?
- Is that the source agency's responsibility?
  - Are SORNs, PIAs, and FRNs enough?
- Is that the researcher's responsibility?
  - From what platform, with what standards?
- Is that a data intermediary's responsibility?
  - Federal Statistical Agencies, FSRDC Network, National Secure Data Service

Next actions, especially for intermediaries

- Get intermediaries to satisfy the minimum requirements for transparency, requiring input from both data suppliers and data users
- Assess the applicability of standards from bioethics to other data domains – do they fit justice, education, workforce, housing, business, social media, etc.?
- Determine ways to highlight examples of research findings and the resulting public benefits, translating this for the public

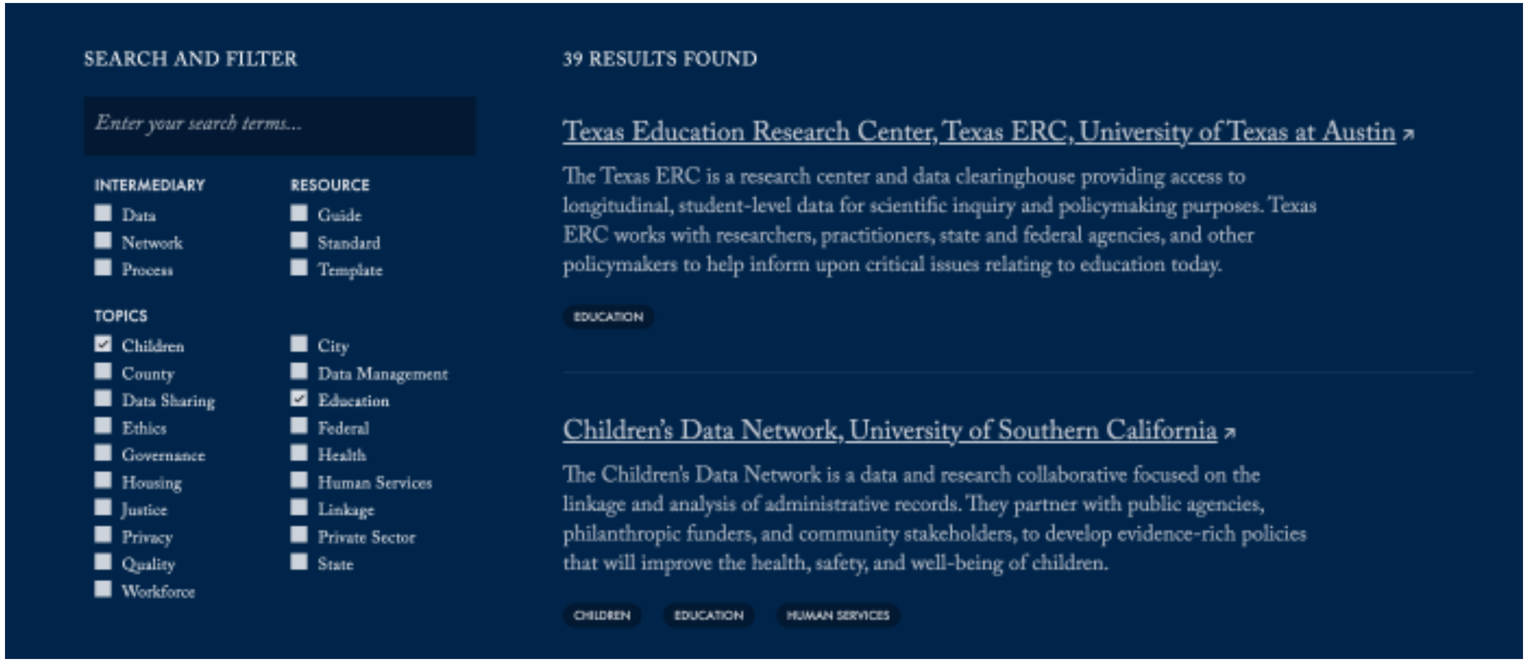
Next actions for ecosystem

- Public awareness to public acceptance
  - Consistent vocabulary
  - Consistent acknowledgements - Program was made possible by the Corporation for Public Broadcasting and by Viewers Like You
- Catalog of harms
  - Possible, probable, perceived
  - No animals were harmed
- Risk of doing nothing
  - What if we don’t link data, evaluate programs, or analyze policies?
  - What if outcomes are unobserved?

amy.ohara@georgetown.edu

Visit our Administrative Data Research Initiative discovery site!

https://adri.georgetown.edu/



SEARCH AND FILTER

Enter your search terms...

INTERMEDIARY
- ☐ Data
- ☐ Network
- ☐ Process

RESOURCE
- ☐ Guide
- ☐ Standard
- ☐ Template

TOPICS
- ☑ Children
- ☐ County
- ☐ Data Sharing
- ☐ Ethics
- ☐ Governance
- ☐ Housing
- ☐ Justice
- ☐ Privacy
- ☐ Quality
- ☐ Workforce
- ☐ City
- ☐ Data Management
- ☑ Education
- ☐ Federal
- ☐ Health
- ☐ Human Services
- ☐ Linkage
- ☐ Private Sector
- ☐ State

39 RESULTS FOUND

Texas Education Research Center, Texas ERC, University of Texas at Austin ↗

The Texas ERC is a research center and data clearinghouse providing access to longitudinal, student-level data for scientific inquiry and policymaking purposes. Texas ERC works with researchers, practitioners, state and federal agencies, and other policymakers to help inform upon critical issues relating to education today.

EDUCATION

Children's Data Network, University of Southern California ↗

The Children's Data Network is a data and research collaborative focused on the linkage and analysis of administrative records. They partner with public agencies, philanthropic funders, and community stakeholders, to develop evidence-rich policies that will improve the health, safety, and well-being of children.

CHILDREN EDUCATION HUMAN SERVICES