



Data,
Policies
and
Algorithms

Denis Parfenov @prfnv

Tokyo Feb 2019

DaMaHub.org



Source:

<https://twitter.com/thereaIbanksy/status/929435067146960896>



“Humankind today faces three common problems that make a mockery of all national borders, and that can only be solved through global cooperation. These are nuclear war, climate change and technological disruption.” - Yuval Noah Harari



*“Prime Minister Shinzo Abe of Japan pledged his country’s G20 chairmanship to the development of an **agenda for global data governance**. Abe said that the process, dubbed the “Osaka track”, would work under the auspices of the World Trade Organization (WTO).”*

Source: <https://www.weforum.org/press/2019/01/world-economic-annual-meeting-2019-closes-with-initiatives-to-address-global-problems/>



Citizens generate data continuously

The video displays a woman walking on a sidewalk, framed by a green bounding box. A red bounding box is also visible on a person crouching nearby. The scene is overlaid with several data visualization elements:

- Profile Card (Top Left):**

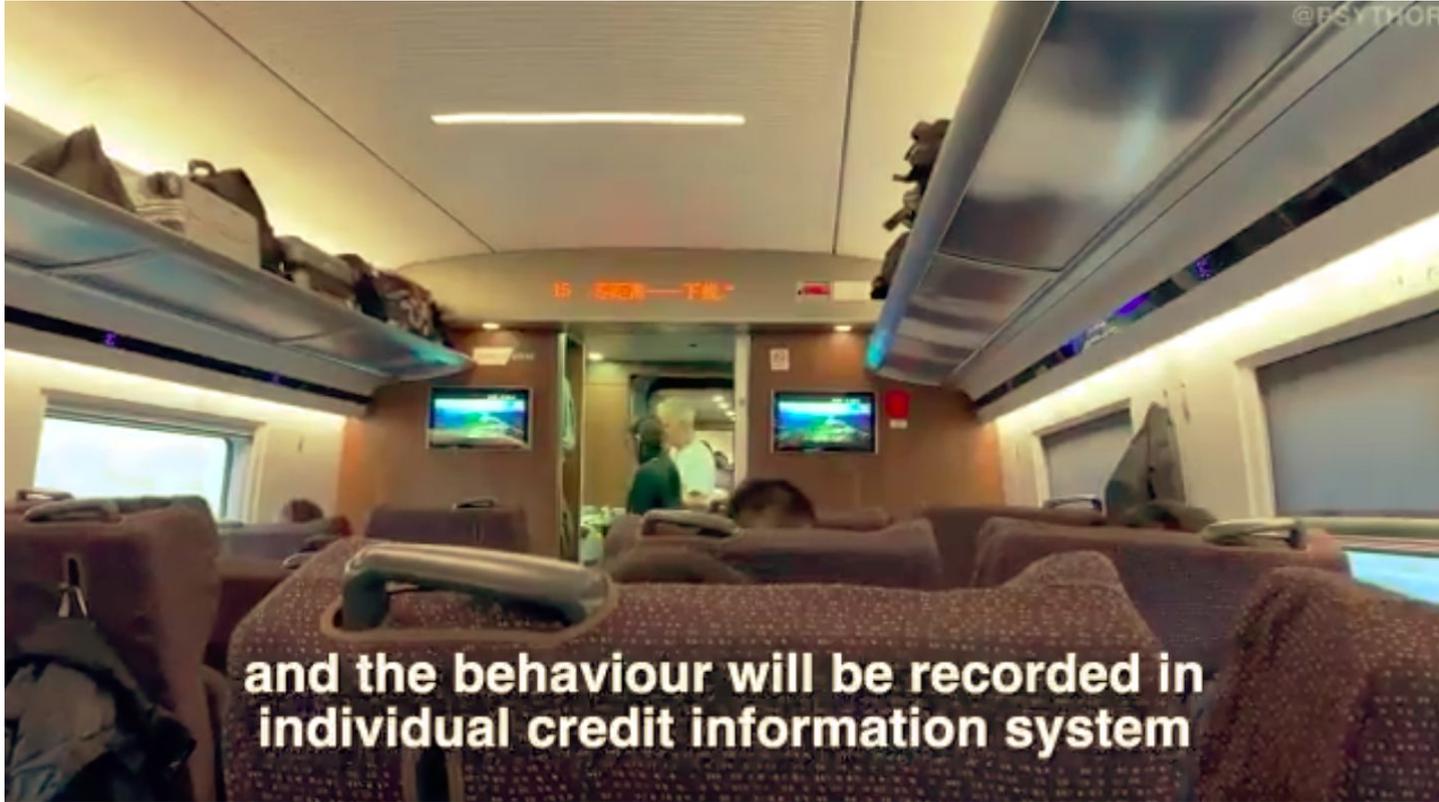
姓名:	范丹丹
性别:	女
年龄:	36
有无犯罪前科:	没有
政治诚信:	诚信
- Social Credit Score (Center):** 社会信用体系 782
- Number List (Bottom Right):**
 - 24.5991825158492
 - 17.2194277610944
 - 9.83967300633967
 - 12.2995912579246
 - 24.5991825158492

The video player interface includes a progress bar at the bottom showing 5:14 / 27:14, and standard playback controls (play, next, volume, settings, full screen, etc.).

Source: https://www.youtube.com/watch?v=eViswN602_k



Algorithms fueled with our data



Source: <https://twitter.com/i/status/1056811593177227264>



Liberty ✓
@libertyhq

Follow

Met Police are aiming #FacialRecognition cameras at #Romford station. They've confirmed to us they will stay here until 6pm, and so will capture the faces of everyone going in and out during rush hour this evening



11:47 AM - 31 Jan 2019

Source: <https://twitter.com/libertyhq/status/1090939629505777665>



GOOGLE CHINA PROTOTYPE LINKS SEARCHES TO PHONE NUMBERS



Ryan Gallagher

September 14 2018, 8:25 p.m.

GOOGLE BUILT A prototype of a censored search engine for China that links users' searches to their personal phone numbers, thus making it easier for the Chinese government to monitor people's queries, The Intercept can reveal.

Source:<https://theintercept.com/2018/09/14/google-china-prototype-links-searches-to-phone-numbers/>



Regulating the internet giants

The world's most valuable resource is no longer oil, but data

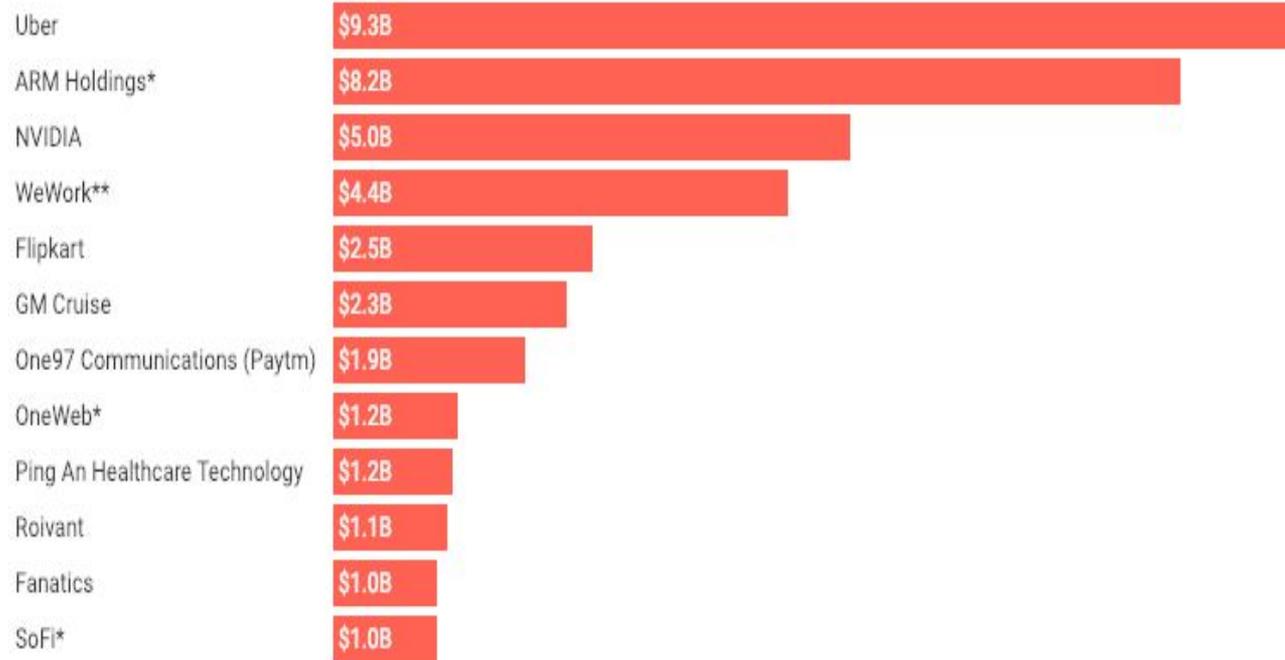
The data economy demands a new approach to antitrust rules



Source: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>



Where SoftBank has invested its \$100 billion Vision Fund?



Source: <https://www.recode.net/2018/2/5/16974032/this-is-where-chart-softbank-vision-fund-masayoshi-son-venture-capital>



Nov 2010 - April 2014

- What can we do to empower the people of Ireland & the govt to make better-informed evidence-based decisions?
- **Let's deploy data-driven open-source applications!**
- **But there is no data!**



Open Gov Partnership 

@opengovpart

Follow



The **#Irish** Experience - Phase 1: The Campaign to Join - See more at: opengovpartnership.org/blog/sarita-ra ...
#Ireland #opengov

8:05 AM - 25 Oct 2013

1 Retweet 2 Likes



XML feed killed Aug 2012

www.kildarestreet.com

our first bookmark by using the heart in the address bar or [Import bookmarks now](#)

KildareStreet

Search GO

Home **Dáil** Seanad Committees Northern Ireland Help

KildareStreet news Recent comments

Join Log in

Hello!

KildareStreet.com is a searchable archive of everything that's been said in the Dáil and all written parliamentary questions since January 2004, everything in the Seanad since September 2002, and all Committee meetings since September 2012.

At KildareStreet.com you can:

- Search:** SEARCH [\(Advanced Search\)](#)
Popular searches today: [public services card](#), [MyGovID](#)
- [Sign up to be emailed when something relevant to you happens in the Oireachtas](#)
- Read and comment on:**
 - Today's Dáil debates:**
 - [Prelude](#)
 - [Ceisteanna - Questions - Ceisteanna ar Sonraíodh Uain Dóibh - Priority Questions](#)
 - [Ceisteanna Eile - Other Questions](#)
 - [Ceisteanna ó Cheannairí - Leaders' Questions](#)
 - [Visit of Malawi Delegation](#)
 - [Ceisteanna ó Cheannairí \(Atógáil\) - Leaders' Questions \(Resumed\)](#)
 - [Ceisteanna ar Reachtáíocht a Gealladh - Questions on Promised Legislation](#)
 - [European Organisation for Astronomical Research in the Southern Hemisphere: Motion](#)
 - [Future of the Post Office Network: Motion \(Resumed\) \[Private Members\]](#)
 - [BusConnects: Motion \(Resumed\) \[Private Members\]](#)
 - [Ábhair Shaincheisteanna Tráthúla - Topical Issue Matters](#)
 - [Central Bank \(National Claims Information Database\) Bill 2018: Order for Second Stage](#)
 - [Central Bank \(National Claims Information Database\) Bill 2018: Second Stage](#)
 - Yesterday's Written Answers:**
 - [Department of Transport, Tourism and Sport](#)
 - [Department of Foreign Affairs and Trade](#)
 - [Department of Finance](#)
 - [Department of Public Expenditure and Reform](#)

? What's all this about?

KildareStreet.com is a non-partisan website which aims to make it easy for people to keep tabs on their elected representatives in the Houses of the Oireachtas.

Follow us

[FACEBOOK](#) [TWITTER](#)

Most recent comments

- Monica Condron:** It is quite astonishing that legislation has not yet been drafted for the Interim TCA. This was announced on the 23rd January 2018 and while there were consultations on the TCA to be introduced... (21 hours, 8 minutes ago)
[Read comment](#) | [All by this user](#)
- Gregory Beck:** Hmm (2 weeks, 3 days ago)
[Read comment](#) | [All by this user](#)
- Benish Arslan:** Hi I have submitted my application on August 2016. application is straight forward no criminal offence no conviction never been anywhere outside Ireland since 7 years but still waiting for... (25 July 2018)
[Read comment](#) | [All by this user](#)
- Paul Enright:** Why is bullying behaviour not dealt with in the public service? Is it because possibly the "ROT" goes a long way up the "pole"... I can understand why so many victims of bullying consider... (23 July 2018)
[Read comment](#) | [All by this user](#)
- N R:** When will this legislation be implemented, as it has been an ongoing discussion since 2015?! At least it would be an eye opener for those who get recruited with false hopes of getting into a... (23 July 2018)
[Read comment](#) | [All by this user](#)

See more comments posted recently

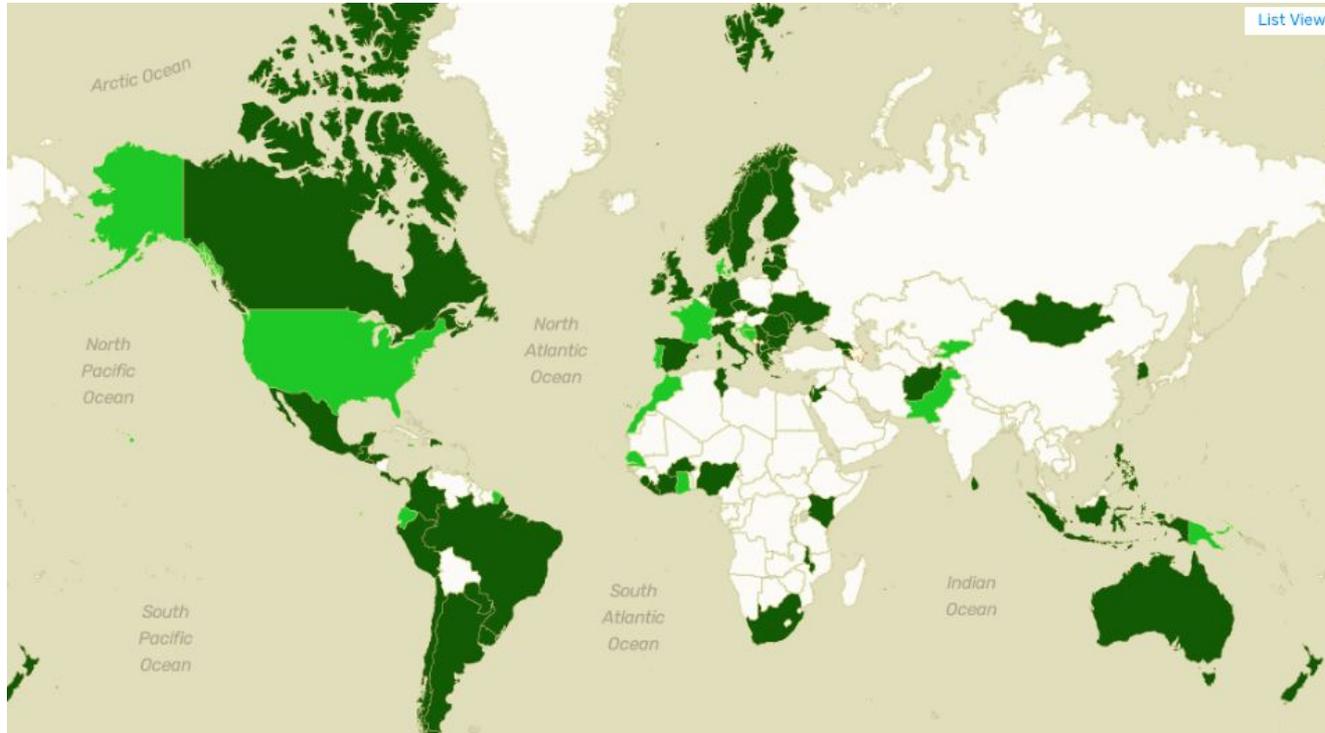


Open Data Ireland Meetups and Hackathons 2012 - 2014



source: ti.to/open-knowledge-ireland/

May 2013 -



Source: <https://www.opengovpartnership.org/countries/ireland>



OPEN
KNOWLEDGE
IRELAND

Open Knowledge Ireland 2013



Source: <http://openknowledge.ie/>



	Fuchi with Mitsudomoe Mon	fuchi	2018-09-07 00:57
	Shishi And Butterfly Tsuba	tsuba	2018-11-15 00:16
	Katchushi Buddhist Wheel Tsuba	tsuba	2018-12-13 01:21
	Large Namban Style Tsuba	tsuba	2018-10-08 00:28
	Jumyō Tanto with simple koshirae	tanto	2013-12-30 01:15



OPEN
KNOWLEDGE
IRELAND

CKAN Hackathon Sept 2013



Source: <http://openknowledge.ie/irish-okfn-chapter-launched-at-ckan-hackathon-in-dublin/>

November 2013

01.1 Establishment of best practice standards for Open Data

01.2 Establishment of Ireland's Open Data Platform

01.5 Establishment of an Open Data Ireland Governance Board (ODIGB) and Steering and Implementation Group (SIG) for Open Data Ireland

01.6 Signing up to the G8 Open Data Charter



OPEN
KNOWLEDGE
IRELAND

Hospital Waiting List(s) 2013 - 2017



Source:<http://openknowledge.ie/projects/open-hospital-waiting-list/>



Hospital Waiting List(s) 2013 - 2017

Data-Information-Knowledge-Action Link (Milton, 2005)

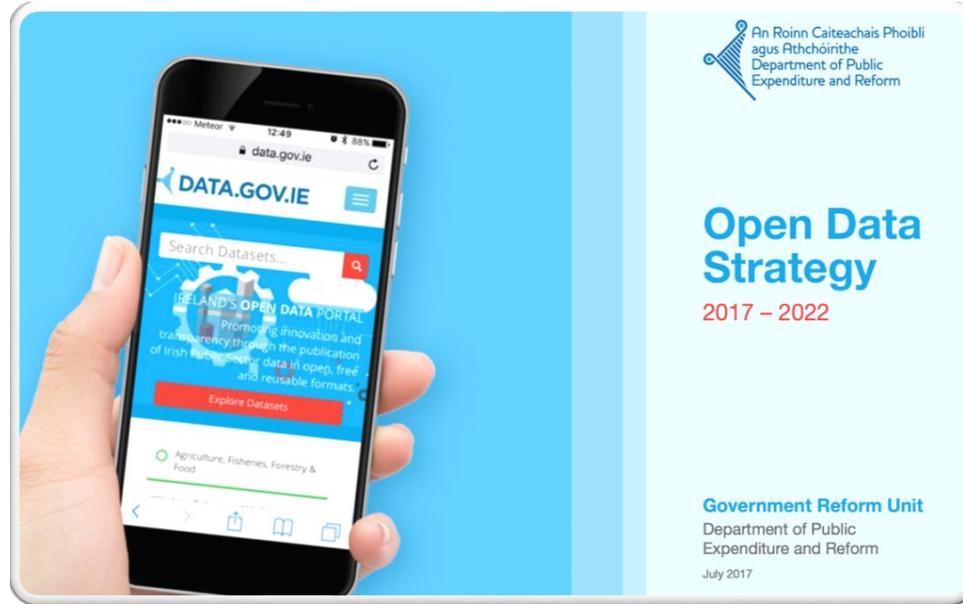


Open Data Governance Board 2015 -



Source: <https://twitter.com/IRLDeptPER/status/669170356775727104>

Open Data Strategy 2017-2022



An Roinn Caiteachais Phoibli
agus Aithchóirithe
Department of Public
Expenditure and Reform

Open Data Strategy

2017 - 2022

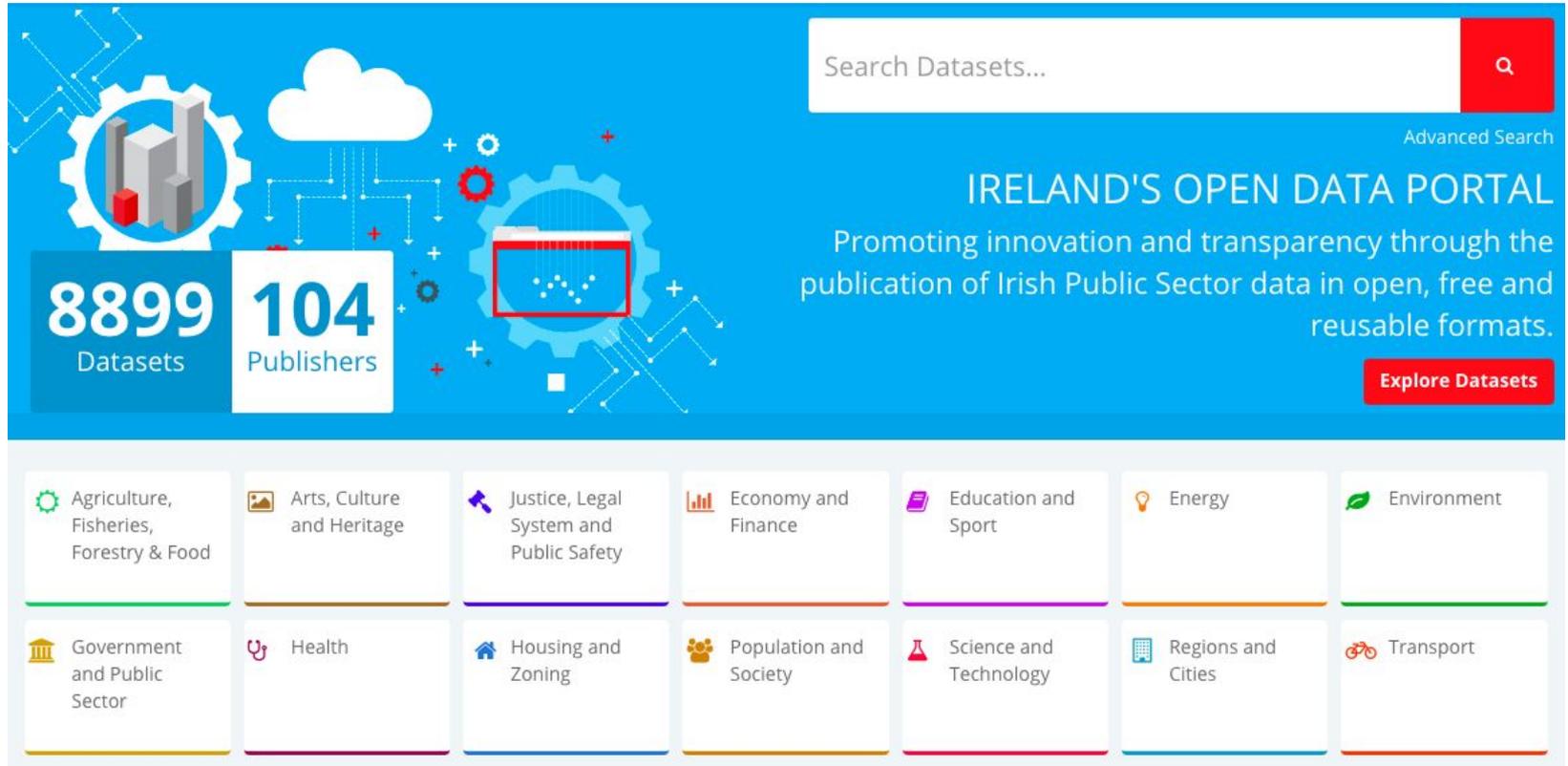
Government Reform Unit

Department of Public
Expenditure and Reform

July 2017

Government Reform Unit
Department of Public
Expenditure and Reform
July 2017

From Zero to 8899 Open Datasets



The screenshot shows the Ireland's Open Data Portal homepage. At the top left, there is a search bar with the text "Search Datasets..." and a red search button. Below the search bar, the text "Advanced Search" is visible. The main heading is "IRELAND'S OPEN DATA PORTAL" in large white letters. Below this, a sub-heading reads "Promoting innovation and transparency through the publication of Irish Public Sector data in open, free and reusable formats." A red button labeled "Explore Datasets" is positioned at the bottom right of this section. On the left side, there are two large white boxes: one containing "8899 Datasets" and another containing "104 Publishers". The background features a blue banner with various icons representing data, technology, and industry. Below the main banner, there is a grid of 14 category tiles, each with an icon and a title. The categories are: Agriculture, Fisheries, Forestry & Food; Arts, Culture and Heritage; Justice, Legal System and Public Safety; Economy and Finance; Education and Sport; Energy; Environment; Government and Public Sector; Health; Housing and Zoning; Population and Society; Science and Technology; Regions and Cities; and Transport.

Search Datasets... 

Advanced Search

IRELAND'S OPEN DATA PORTAL

Promoting innovation and transparency through the publication of Irish Public Sector data in open, free and reusable formats.

[Explore Datasets](#)

8899 Datasets

104 Publishers

-  Agriculture, Fisheries, Forestry & Food
-  Arts, Culture and Heritage
-  Justice, Legal System and Public Safety
-  Economy and Finance
-  Education and Sport
-  Energy
-  Environment
-  Government and Public Sector
-  Health
-  Housing and Zoning
-  Population and Society
-  Science and Technology
-  Regions and Cities
-  Transport

Open Data (Re-)Use

14 showcases found

Order by: Relevance



Irish Rail - Realtime

A showcase webpage using a realtime API from Irish Rail showing the live position (data refreshed every minute) of trains across Ireland. Using another API query train stations...

1 Dataset



Social Housing Units Completed

Social housing units completed

1 Dataset



Challenge: Funding



John Handelaar @handelaar · 21 Dec 2018

Downside: switching it off means that I've got to rewrite *all* the parsers AGAIN by January 14th.

Pass it on? Expired cards mean I now do all this for less than **€40** a week

If you used to donate but the card ran out, or to support like you did in 2012?

kldr.st/FiverAMonth

kildarestreet.com @KildareStreet

Let it be recorded for posterity that as of today, the web publishing project commissioned and paid for by spiteful Oireachtas staff with your money in 2012, *designed* to prevent us from operating, and which has *never* even worked properly... is decommissioned.

Show this thread





Challenge: Persistent Access





Research Data Workflow

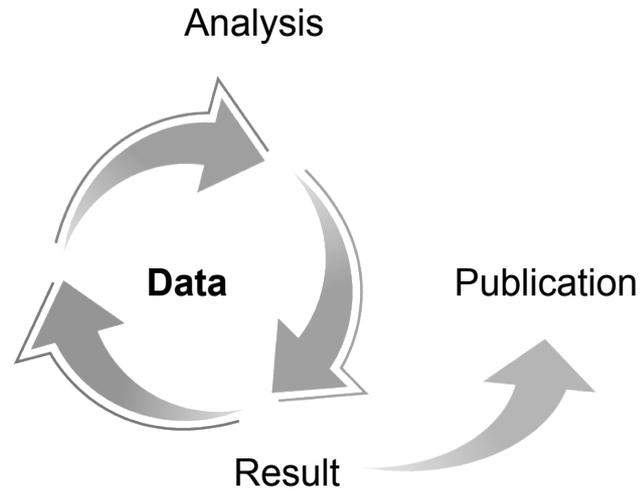
Idea + Funding



Experiment



Preprocessing





"A record, if it is to be useful to science, must be continuously extended, it must be stored, and above all it must be consulted. "

-- Vannevar Bush
"As We May Think"
1945



Problems with Academic Publishing Today

- 30,000+ Academic Journals
- 3,200+ Institutional Repositories (IR) launched since 1999
 - Interoperability of data remains a major issue



What about Research Data?

- 90% of scientific data generated in the last 2 years
- Only 12% of NIH-funded data makes it to data repositories
 - Report of HLEG on European Open Science Cloud, 2016
- Requesting data from the author and journal supplementary data are insufficient

RESEARCH ARTICLE

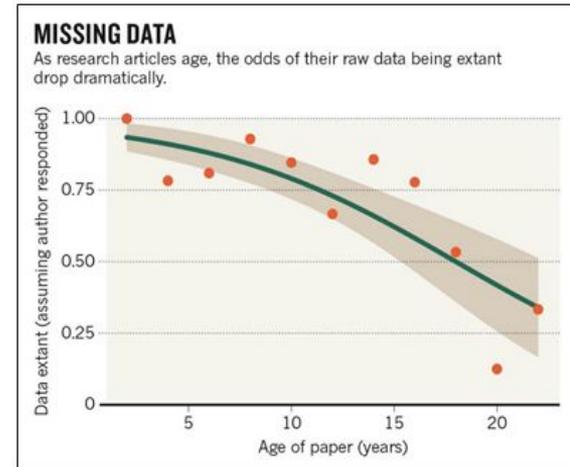
Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot

Martin Klein, Herbert Van de Sompel, Robert Sanderson, Harihar Shankar, Lyudmila Balakireva, Ke Zhou, Richard Tobin

RESEARCH ARTICLE

Scholarly Context Adrift: Three out of Four URI References Lead to Changed Content

Shawn M. Jones, Herbert Van de Sompel, Harihar Shankar, Martin Klein, Richard Tobin, Claire Grover



Vines, T. H. et al. doi: 10.1016/j.cub.2013.11.014



“3 years after finishing the project **80%** of all researchers can **not present the data** anymore.”

Prof. Dr. Mark Ferguson, SFI, 2016



 **Margaret Hannum** @Margaret_Hannum · 23h
Extra disappointing when it's juuuust the code you think you'd need 🙄
[#reproducibility](#) [#rstats](#) [#openscience](#)

**WHEN A PAPER SAYS THEY PROVIDE ALL DATA
AND CODE FOR REPRODUCIBILITY BUT THE LINK IS BROKEN**



All right, then. Keep your secrets.

🗨️ ↻ 6 ❤️ 32 ✉️ 📌

Source: https://twitter.com/Margaret_Hannum/status/1092182263155048449



Link / Reference Rot



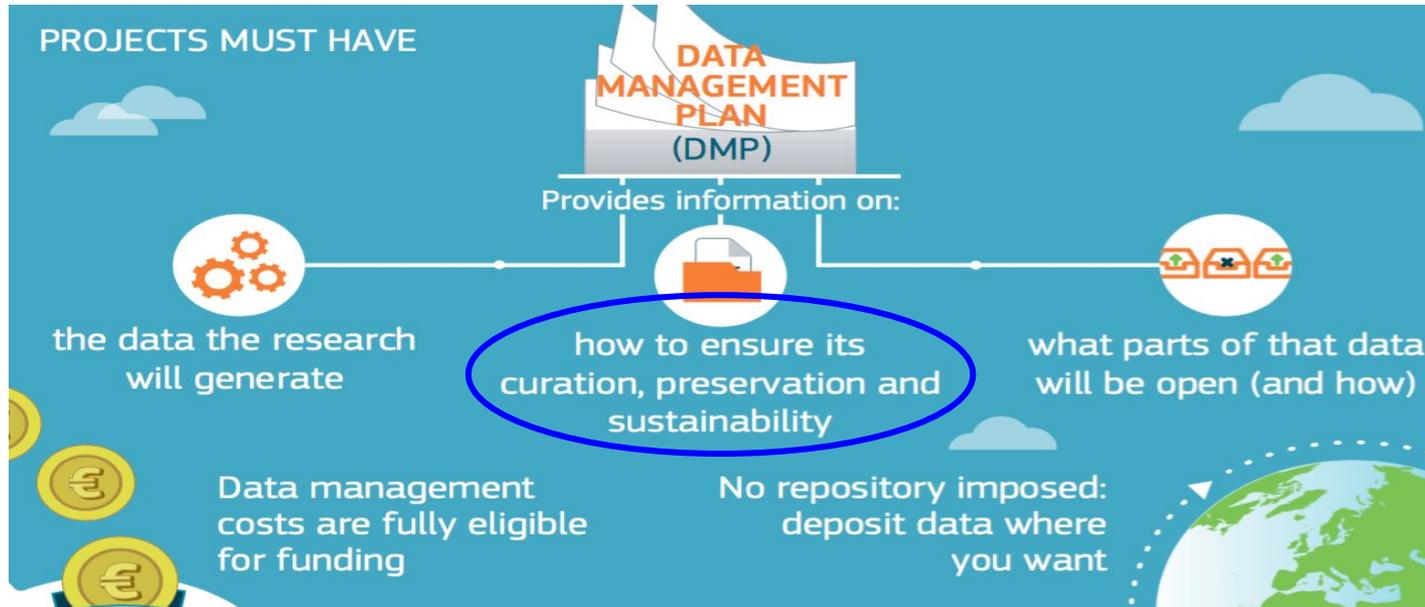
- 70% of the URLs within Harvard Law Review (HLR), the Harvard Journal of Law and Technology (JOLT) and the Harvard Human Rights Journal (HRJ) suffer **link rot**
- 50% of the URLs within U.S. Supreme Court opinions suffer **reference rot**



Source: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2329161



Since 2016 Data Management Plan mandatory

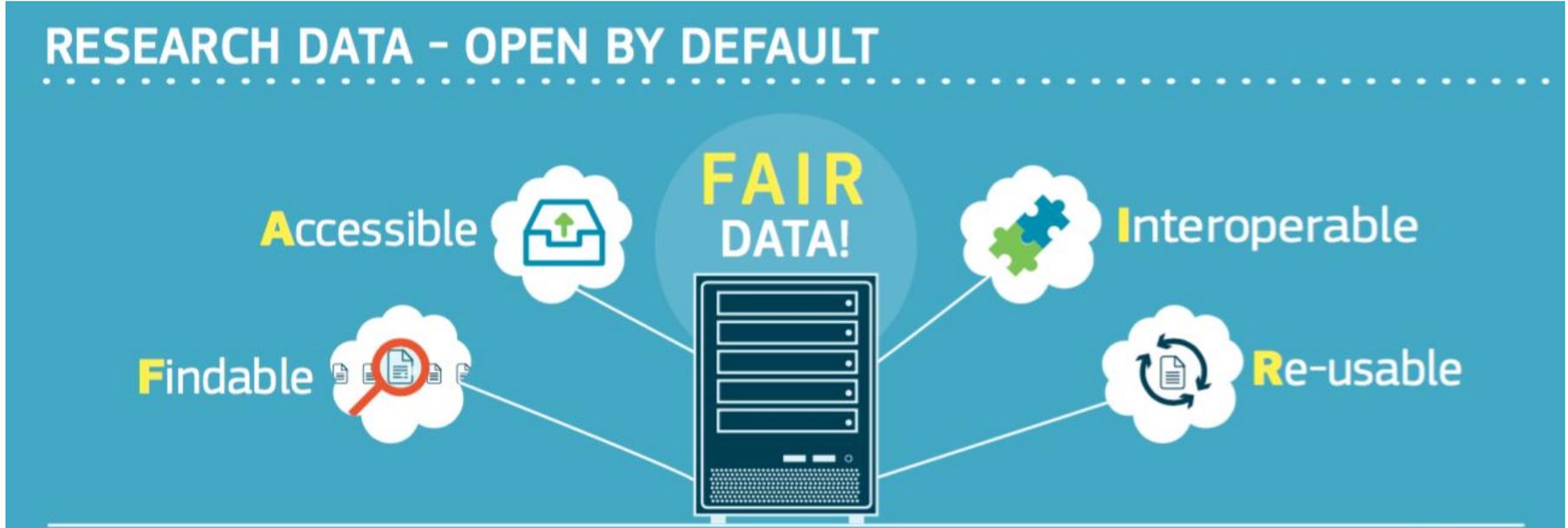


❑ Data Management plan and preservation of scientific outcomes is mandatory now:

EU: Horizon 2020; **Germany:** BMBF, DFG, DAAD, HRK, Alexander von Humboldt-Stiftung, Deutsche Akademie der Naturforscher Leopoldina, Fraunhofer-Gesellschaft, Max-Planck-Gesellschaft Wissenschaftsrat; **UK:** AHRC, BBSRC, CRUK, EPSRC, ESRC, MRC, NERC, STFC, Wellcome Trust; **And many, many more...**



Publicly funded research data must remain FAIR





Digital Single Market: EU negotiators agree on new rules for sharing of public sector data

Brussels, 22 January 2019

“Member States will be required to develop policies for open access to publicly funded research data while harmonised rules on re-use will be applied to all publicly-funded research data which is made accessible via repositories.”



Data citation and versioning

“The demand for reproducibility of research results is growing. [There is need] to **reference the exact version of the data** that was used to underpin the research findings, and/or was used to generate higher level products. ”

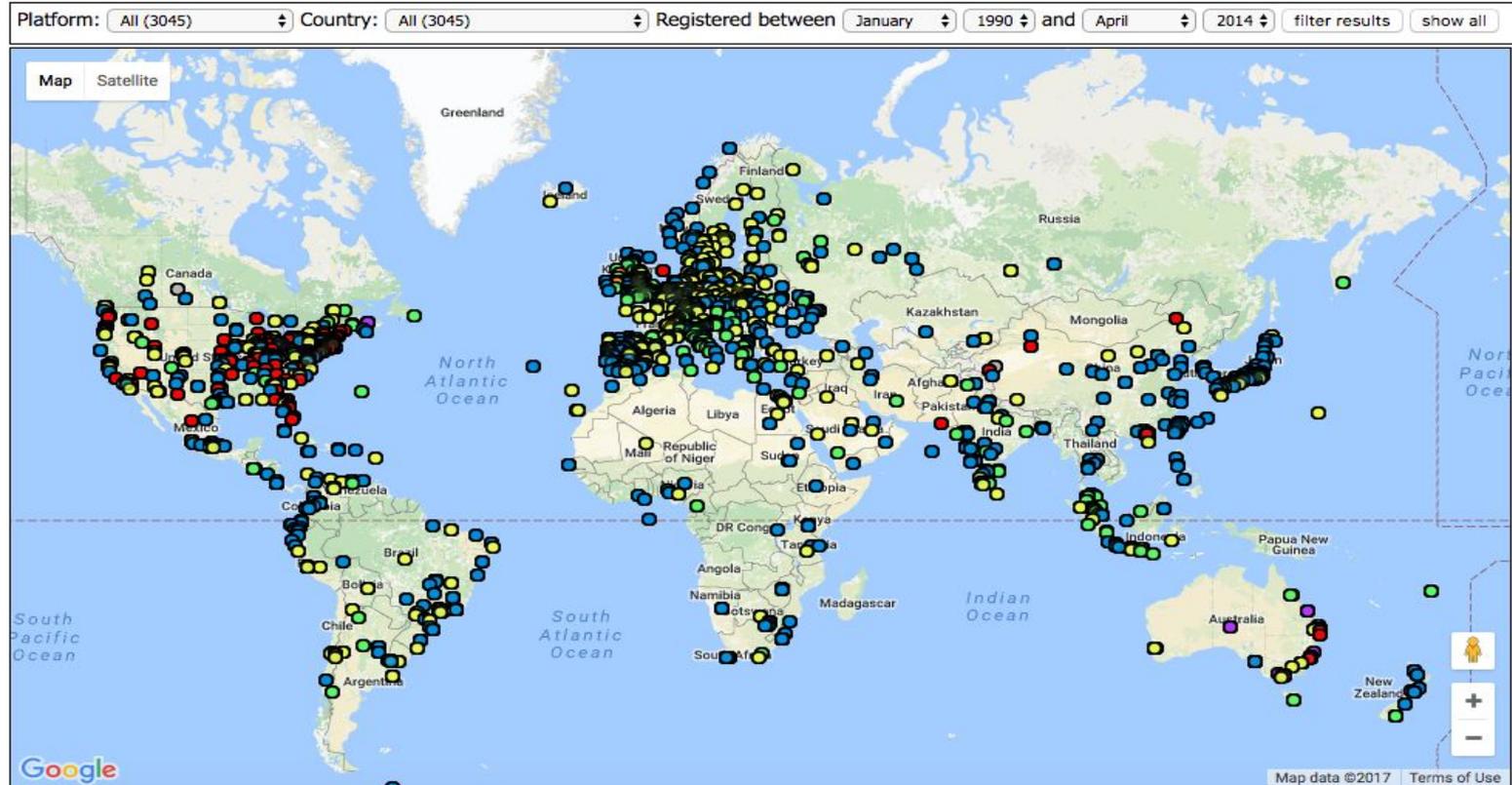


Data Citation WG

Data Versioning WG



The existing digital infrastructure for preservation of data needs to be complemented



Source: <http://maps.repository66.org>



Data Management Hub



simple



social



neutral



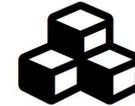
damahub.org



secure



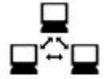
authentic



resistant



What researchers need?



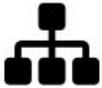
Distributed, neutral



Timestamps for
Authenticity



Easy Sync

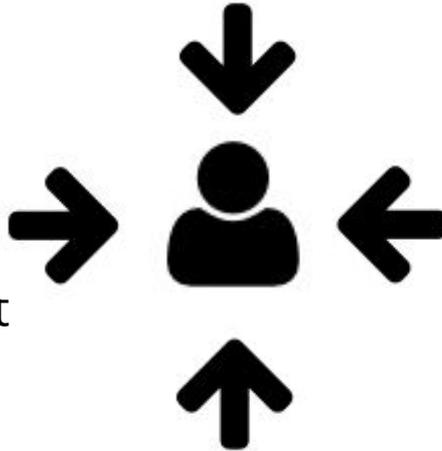


Version-Management



Private Data

**Data
Authorship
Attribution**



Peer-Review,
Comments



Open Access



Open Source



No Vendor-Lockin



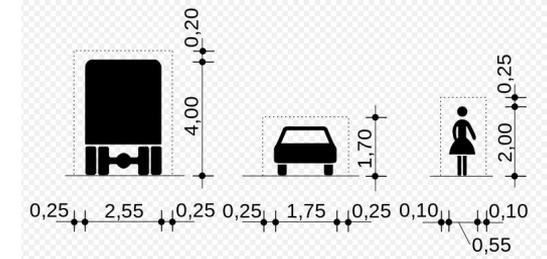
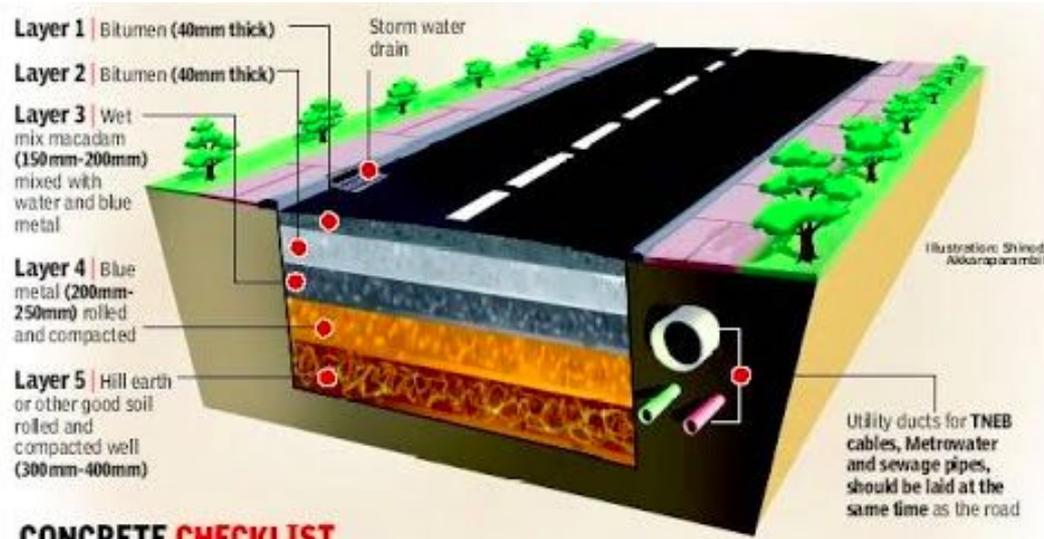
**Lots Of Copies Keep
Stuff Save**

Roman Road Infrastructure





Roads Standards and Structure



CONCRETE CHECKLIST

▶ Road traffic, pavement, parking space and maneuverability for big vehicles such as bus/truck should be considered

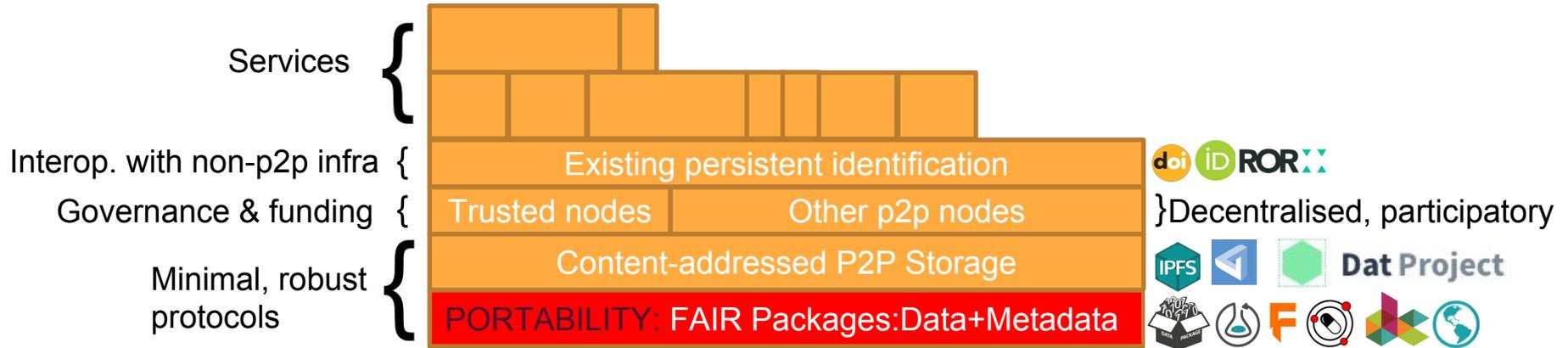
▶ Middle portion of the road should have a gradient that is about 2.5% higher than the rest of the surface so that rainwater flows into the stormwater drain

▶ Keeping in mind the rainfall intensity in the city, storm water drains should be built as per best standards of hydraulics

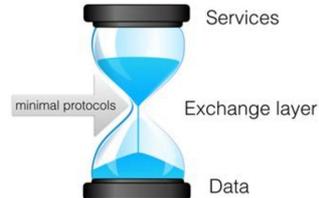
▶ A stormwater channel should have an inlet to capture runoff quickly. The current 4-inch pipe on the roadside clogs easily



Data-Road Structure and Standards



“An Internet of FAIR data and services”



Source: @eocarragain



Static, portable “Data Packages”



BagIt



image: @OA_RHUL



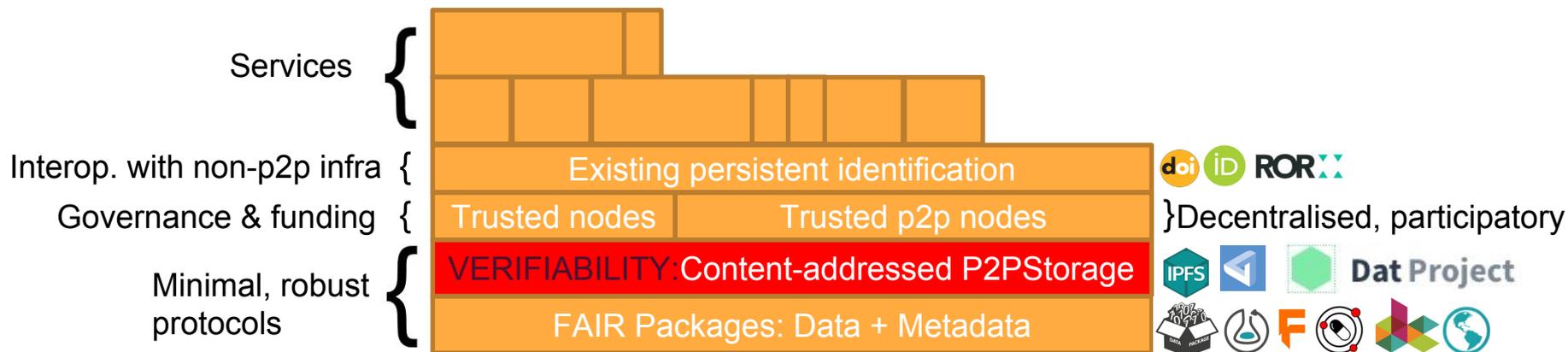
DwC Archive



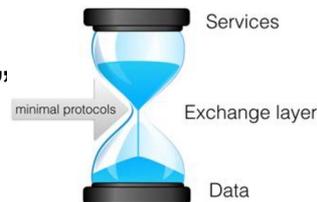
“Make **Data Crate** again!”



Data-Road Structure and Standards



“An Internet of FAIR data and services”

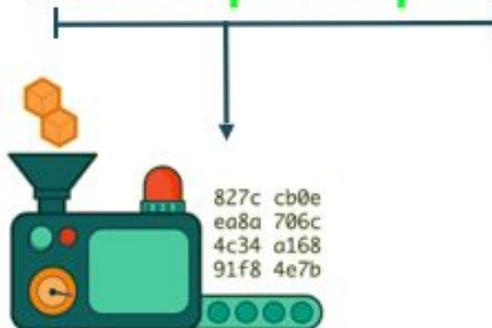


Source: @eocarragain



LOCATION-ADDRESSING

<http://site.com/data/pids.pdf>



<ipfs://zdj7WjqNriReTcEveRh/pids.pdf>

CONTENT-ADDRESSING

Image: [@procollabs](#)

Source: <https://ipfs.io/ipfs/QmdhEQrRvvo6mY4iyNz544xChM6nmrJbefvnYNEtVTQzz5/CATALOG.html>



**OPEN
KNOWLEDGE
IRELAND**

HWL Data on IPFS (2016)

Hospital Waiting List for Dec 2015 is now available on IPFS.

A) **okfn-ie-hospital-data**

<https://ipfs.io/ipfs/QmT66oHDwzb8dU5vnZt3Ez5aStcWCjbqjNE2pA25ShTjmM/>

(B) **hospital dashboard**

<http://openknowledge.ie/>

(Dan Alexandru Bujoreanu, 2015)

IPFS

Index of /ipfs/QmT66oHDwzb8dU5vnZt3Ez5aStcWCjbqjNE2pA25ShTjmM/

..	
okfn-ie-hospital-data	

NTPF_Datastore - Fact_Inpatient.csv

NTPF_Datastore - Fact_National.csv

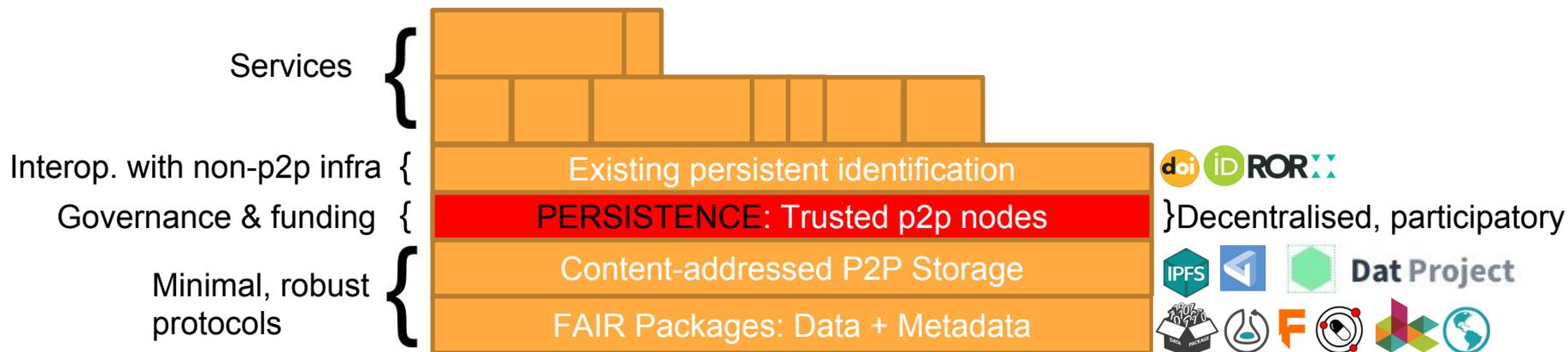
NTPF_Datastore - Fact_Outpatient.csv

NTPF_Datastore - Linked_Hospital.csv

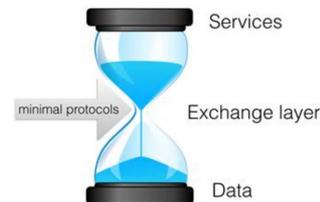
Source:<http://openknowledge.ie/projects/open-hospital-waiting-list/>



Data-Road Structure and Standards



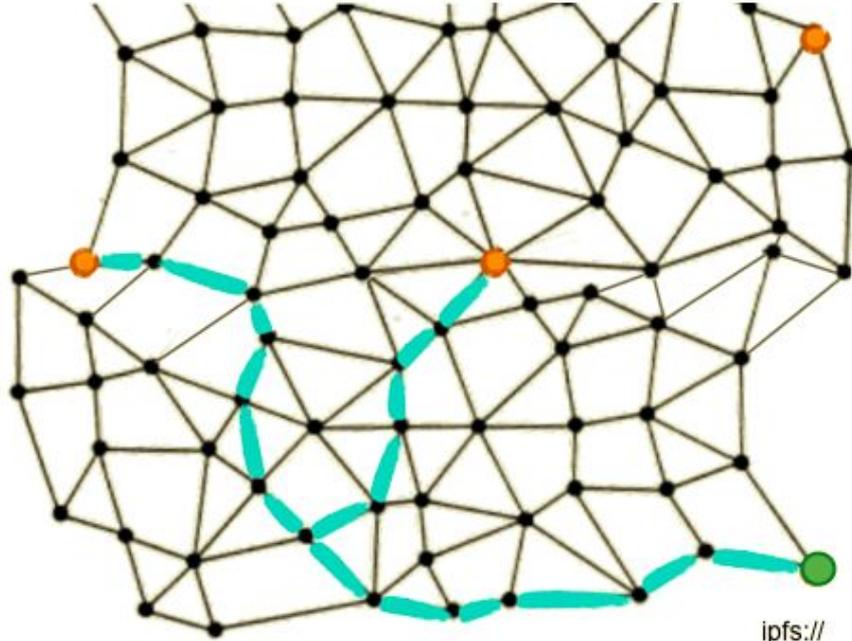
“An Internet of FAIR data and services”



Source: @eocarragain



Co-ordinated, well governed decentralisation!



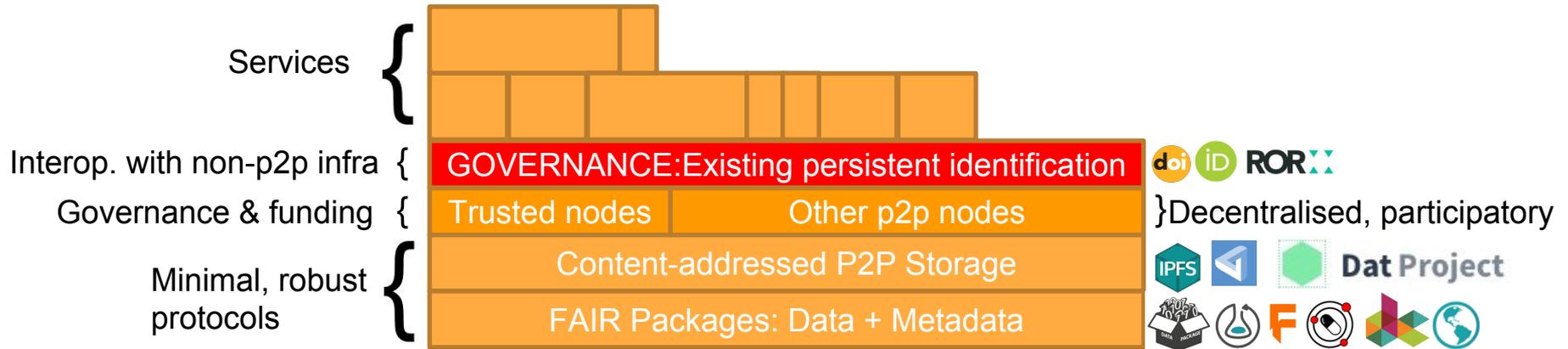
ipfs://
zdj7WjqNrjReTcEveRhgcXs
JvSGwLxJ7js1R7ZCzNaQSK
uTh

image: [@flyingzumwalt](#)

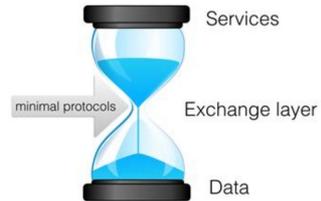
Source: [@eocarragain](#)



Data-Road Structure and Standards



“An Internet of FAIR data and services”



Source: @eocarragain

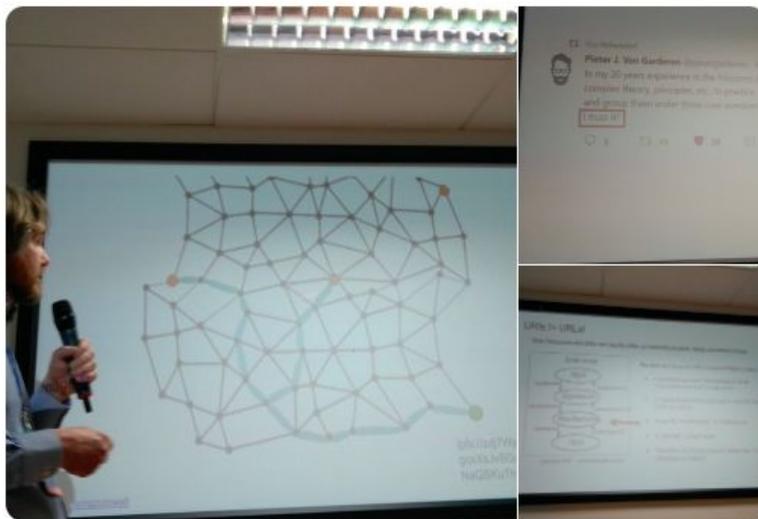


Data M@agement Hub

@DaMaHub



. @eocarragain presents distributed technologies for preservation of content-addressed data. @UCCLibrary #PIDapalooza19 #PID19 #damahub #dweb #ipfs



3:42 AM - 23 Jan 2019

Source: <https://twitter.com/DaMaHub/status/108803926560128614>



 You Retweeted



Pieter J. Van Garderen @pjvangarderen · Apr 11



In my 20 years experience in the #digipres domain I have read a fair share of complex theory, principles, etc.. In practice, I am always able to simplify 'things' and group them under three core questions: 1) can I find it? 2) can I use it? 3) can I trust it?





DaMaHub: Find Data, Use Data, Trust Data



+



+



+



Portable
research data
packaging
specifications:
DATA +
METADATA

Interoperable,
platform/
technology
agnostic

A large-scale
cluster of
trusted,
persistent
IPFS nodes

Centrally
co-ordinated,
funded,
governed, but
3rd parties can
replicate

Coordinate
what should
be stored and
by who in a
verifiable way

Verify/Audit
that the
“trusted
cluster” is
doing its job

Index,
discovery/search
engine

Because you
need to find stuff



Data M@nagement Hub @DaMaHub · 22 Apr 2018



Many thanks to everyone who participated in discussing existing opportunities and laying out foundation for development of decentralised data infrastructure for science.



You, The British Library and IPFS



Source: <https://twitter.com/DaMaHub/status/988073832668901376>

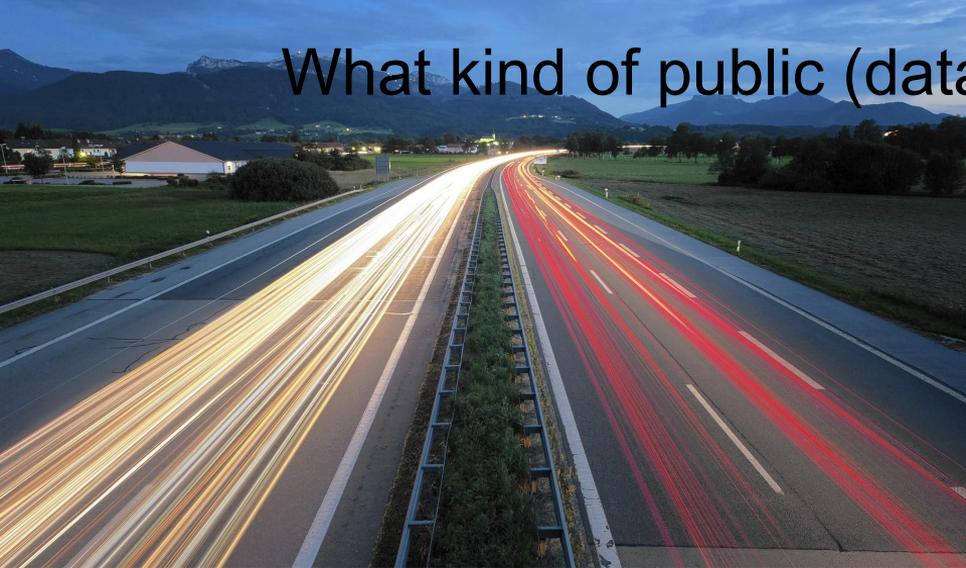


"Never doubt that a small group of thoughtful, committed, citizens can change the world. Indeed, it is the only thing that ever has."

– Margaret Mead

But resources required....

What kind of public (data)-roads will we have?





Next

- [“Steps”](#) - working retreat in Lithuania: 5-14 July 2019  **BLOCKCHAIN**
for Science
- [“DWEB for Science”](#) - symposium in London: April 2020 



検証可能なデータに信頼する



Data Management Hub

DaMaHub.org

Github.com/DaMaHub

Gitter.im/DaMaHub

Twitter.com/DaMaHub

Data Management Hub

Denis Parfenov [@prfnv](#)

Eoghan Ó Carragáin [@eocarragain](#)

James Littlejohn [@aboynejames](#)

データ管理ハブ

References

ActiveCitizen

- <https://www.scribd.com/user/141976247/ActiveCitizen>
- https://twitter.com/ActiveCitizen_/status/708297480774029316

OGP Ireland

- <https://www.opengovpartnership.org/stories/irish-experience-phase-1-2015-2017-join#sthash.mAUDU2Wh.dpuf>
- <https://opengovpartnership.org/country/ireland>

Open Knowledge Ireland

- <https://ti.to/open-knowledge-ireland>
- <http://openknowledge.ie>

Open Data Governance Board

- <https://data.gov.ie/blog/inauguralmeetingoftheopendatagovernanceboardireland>
- <https://data.gov.ie/>

Data Management Hub

- <http://damahub.org/>
- <https://github.com/damahub>
- <https://ti.to/damahub>



OPEN
KNOWLEDGE
IRELAND



DaMaHub



BLOCKCHAIN
for Science