

Towards a data driven ICANN

Jay Daley, Jonathan Zuck, Ed Lewis
ICANN Copenhagen 2017

Why this session?

- Multiple data requirements in current work
 - gTLD Marketplace Health Indicators
 - Competition, Consumer Choice, Consumer Trust
 - Commercial Business Users Constituency
- ICANN responding
 - Open Data Initiative
- Many registries working on data
 - Domain name classification
 - DNS Popularity
 - Threat detection
 - Name similarity
- !! No overall framework around this !!



Why data matters

- Evidence based policy
- Organisational/community development
- Cleaner and safer DNS
- Business - more, new, better
- Societal impact of DNS



Format of this session

- Examine each of these six topics in
 - Evidence based policy
 - Organisational/community development
 - Cleaner and safer DNS
 - Business - more, new, better
 - Societal impact of DNS
 - Putting this into practice
- Presentation followed by open discussion
 - Roaming microphones
 - Big conversation
 - Harness and direct community energy



Evidence based policy



Current use of data in policy

- Already some excellent examples
IANA SLE approximations by Marc Blanchet
- But data is not public and so ...
No reproducibility
Agenda set by those who pay the analysts
Very low throughput of research
- Some people have their own data
Gives them real power in the debate
e.g. Root server operators and WPAD debate
“data is the new oil”
- So much work going on, how do we cope?

CCT Review draft report

- 30 out of 50 recommendations are data related
 - “Collect wholesale pricing for legacy gTLDs”
 - “Collect TLD sales at a country-by-country level”
- Main recommendation on data
 - “ICANN should establish a formal initiative, perhaps including a dedicated data scientist, to facilitate quantitative analysis, by staff, contractors and the community, of the domain name market and, where possible, the outcomes of policy implementation. This department should be directed and empowered to identify and either collect or acquire datasets relevant to the objectives set out in strategic plans, and analysis and recommendations coming from review teams and working groups.”

Commercial business users

- Wrote letter to the ICANN CEO
“The Commercial Business Users Constituency, the Intellectual Property Constituency, and the Internet Service Providers Constituency are writing to request that ICANN make the collection and publication of data a priority, and that the Board and CEO commit to expeditiously providing the public with unfettered, routine access to raw, unfiltered data related to ICANN’s mission.”



gTLD marketplace health

- Three areas

Robust competition

Marketplace stability

Trust

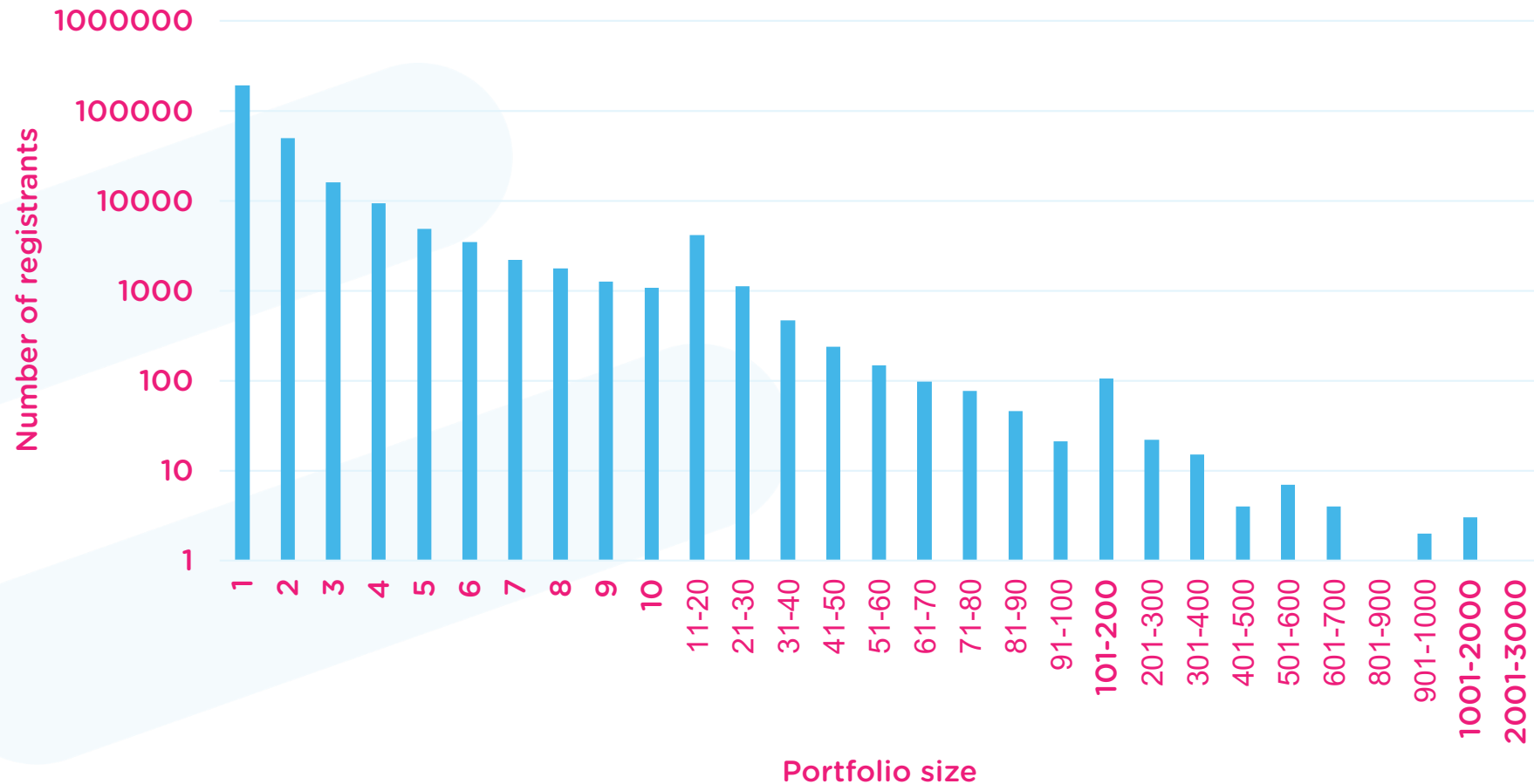
- Indicators rely on data for evidence

What languages/scripts are supported by market

Market concentration segmented in multiple ways

Case study - hoarding?

Number of registrants by portfolio size





Challenge

- Identify datasets
- Make accessible and usable
- Meet demand and support innovation
- Build skills within community
- Culture of sharing and examination of evidence
- Wrap data governance around this
- **Open Data means Open Debate**



Organisational development



The open organisation

“Sunlight is the best disinfectant”

- Applied to
Diversity, remuneration, expenses, etc
- Significant organisational benefits
Reinforces best behaviour
Create culture of community/customer audit
- Until now this has been community led
Case study - travel funding
Case study - ICANN diversity

Case study – Travel funding

- ICANN funds travel for some attendees
- Until recently this data was hard to use
 - Only published in PDFs
 - Some meetings missing
 - Names spelled inconsistently
 - No summaries or multi-meeting analysis
- At Dublin meeting spent several hours
 - Extracting data from PDFs
 - Tidying up names
 - Creating reports
 - Publishing data

Example PDF

ICANN 51st International Public Meeting Los Angeles, United States 12 - 16 October 2014

Report on Travel Support for At-Large Members
US Dollars - Unaudited as of 12/1/2014

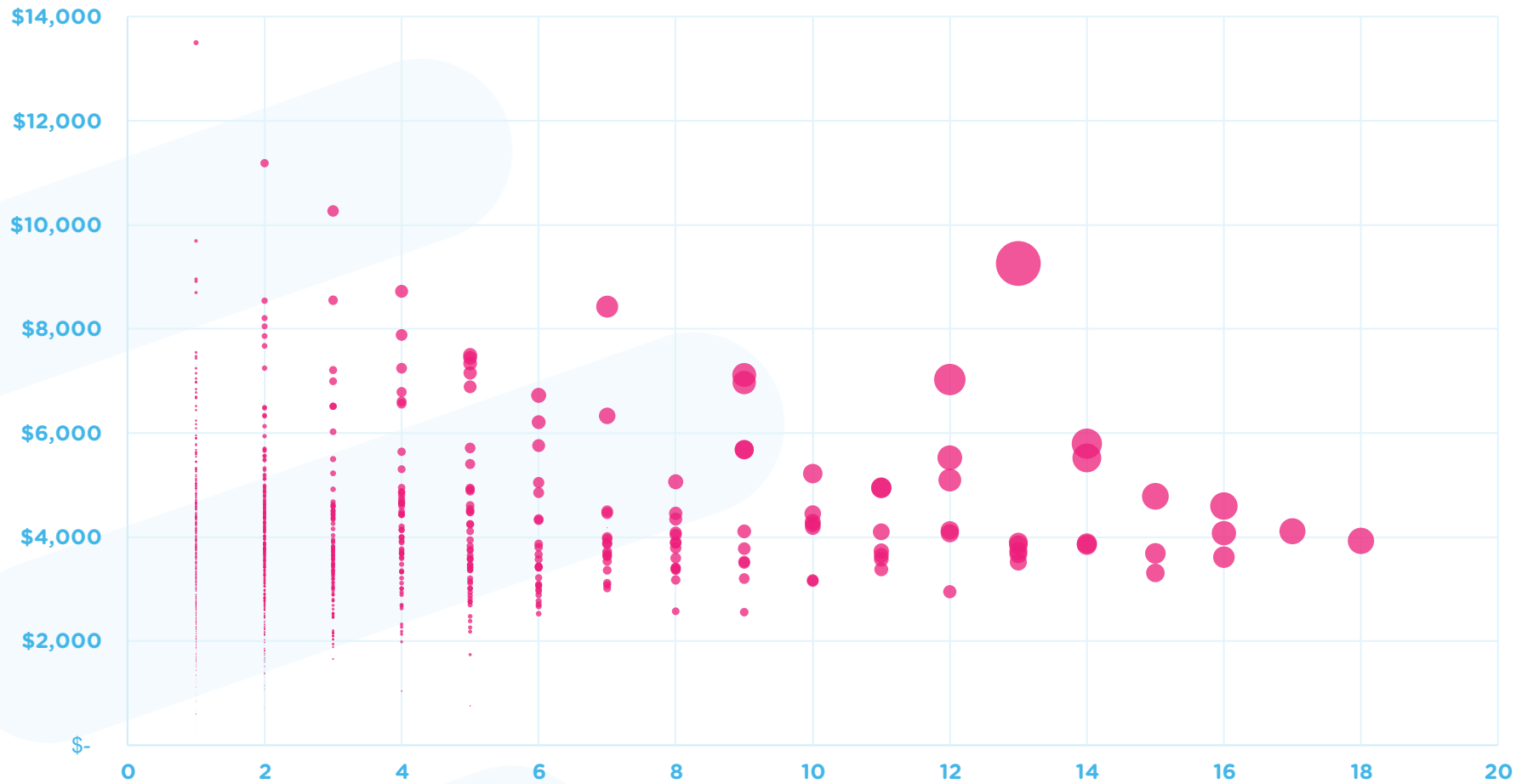
	First Name	Last Name	Group	Class of Travel	Airfare	Hotel ^f	Per Diem /Stipend	Visas	Other	TOTAL	Full Support Equivalent
1	Ali	Ali Alwai Almeshal	ALAC	Economy	\$ 1,723.20	\$ 1,250.00	\$ 400.00	\$ 160.00	\$ -	\$ 3,533.20	1
2	Mohamed Tijani	Ben Jemaa	ALAC	Economy	\$ 1,313.30	\$ 1,500.00	\$ 450.00	\$ 33.00	\$ -	\$ 3,296.30	1
3	Fatima	Cambronero	ALAC	Economy	\$ 2,187.23	\$ 1,250.00	\$ 400.00	\$ 160.00	\$ -	\$ 3,997.23	1
4	Olivier	Crépin-Leblond	ALAC	Business	\$ 4,271.70	\$ 1,500.00	\$ 450.00	\$ -	\$ -	\$ 6,221.70	1
5	Eduardo	Diaz	ALAC	Economy	\$ 850.00	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 2,500.00	1
6	Mohamed	El Bashir Ahmed Mohar	ALAC	Economy	\$ 1,461.90	\$ 1,000.00	\$ 400.00	\$ 179.00	\$ -	\$ 3,040.90	1
7	Rafid	Fatani	ALAC	Economy	\$ 1,244.60	\$ 1,250.00	\$ 350.00	\$ 160.00	\$ -	\$ 3,004.60	1
8	Alan	Greenberg	ALAC	Economy	\$ 562.75	\$ 1,750.00	\$ 400.00	\$ -	\$ -	\$ 2,712.75	1
9	Abdelaziz	Hilali	ALAC	Economy	\$ 2,156.60	\$ 1,500.00	\$ 450.00	\$ -	\$ -	\$ 4,106.60	1
10	Maureen	Hilyard	ALAC	Economy	\$ 2,480.60	\$ 1,500.00	\$ 400.00	\$ -	\$ -	\$ 4,380.60	1
11	Sandra	Hoferichter	ALAC	Economy	\$ 1,205.80	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 2,855.80	1
12	Evan	Leibovitch	ALAC	Economy	\$ 703.55	\$ 1,500.00	\$ 350.00	\$ -	\$ -	\$ 2,553.55	1
13	Wolfgang	Ludwig	ALAC	Business	\$ 4,239.60	\$ 1,500.00	\$ 400.00	\$ -	\$ -	\$ 6,139.60	1
14	Glenn	McKnight	ALAC	Economy	\$ 778.85	\$ 1,250.00	\$ 300.00	\$ -	\$ -	\$ 2,328.85	1
15	Yuliya	Morenets	ALAC	Business	\$ 6,089.50	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 7,739.50	1
16	Gordon Paul Anth	Niigani	ALAC	Economy	\$ 846.35	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 2,496.35	1
17	Barrack	Otieno	ALAC	Economy	\$ 2,270.20	\$ 1,500.00	\$ 400.00	\$ -	\$ -	\$ 4,170.20	1
18	Hadja Fatimata	Ouattara Sanon	ALAC	Economy	\$ 2,398.60	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 4,048.60	1
19	Harriet	Raiche	ALAC	Economy	\$ 2,150.00	\$ 1,500.00	\$ 450.00	\$ -	\$ -	\$ 4,100.00	1
20	Leon Felipe	Sanchez Ambia	ALAC	Economy	\$ 412.25	\$ 1,500.00	\$ 350.00	\$ -	\$ -	\$ 2,262.25	1
21	Vanda	Scartezini	ALAC	Economy	\$ 2,412.84	\$ 1,500.00	\$ 400.00	\$ -	\$ -	\$ 4,312.84	1
22	Jimmy	Schulz	ALAC	Economy	\$ 1,545.30	\$ 1,250.00	\$ 400.00	\$ -	\$ -	\$ 3,195.30	1
23	Ramon Julio Alber	Soto	ALAC	Business	\$ 5,141.83	\$ 1,500.00	\$ 400.00	\$ -	\$ -	\$ 7,041.83	1
24	Jean-Jacques	Subrenat	ALAC	Economy	\$ 1,424.80	\$ 1,500.00	\$ 450.00	\$ -	\$ -	\$ 3,374.80	1
25	Dev Anand	Teelucksingh	ALAC	Economy	\$ 923.30	\$ 1,500.00	\$ 450.00	\$ -	\$ -	\$ 2,873.30	1
26	Siranush	Vardanyan	ALAC	Economy	\$ 1,620.90	\$ 1,250.00	\$ 400.00	\$ 160.00	\$ -	\$ 3,430.90	1
DROPS											
1	Garth	Bruen	ALAC		\$ 424.20	\$ -	\$ -	\$ -	\$ -	\$ 424.20	
2	Humberto	Carrasco Blanc	ALAC		\$ 1,117.20	\$ -	\$ -	\$ -	\$ -	\$ 1,117.20	
3	Philip	Johnson	ALAC		\$ -	\$ -	\$ -	\$ -	\$ -	\$ -	
TOTAL:					\$ 53,956.95	\$ 36,000.00	\$ 10,450.00	\$ 852.00	\$ -	\$ 101,258.95	26

NOTICE: Where actual hotel, flight, and other expenses are not yet available they have been estimated based on number of authorized nights.

Full Support Equivalent: Some travelers only receive airfare, while other received only lodging and per diem; partially supported travelers appear as partial support.

Example output

Average travel funds received by Number of meetings attended





Case study – Diversity

- Report from AFNIC on ICANN Diversity
- Based on public data – but not easy!

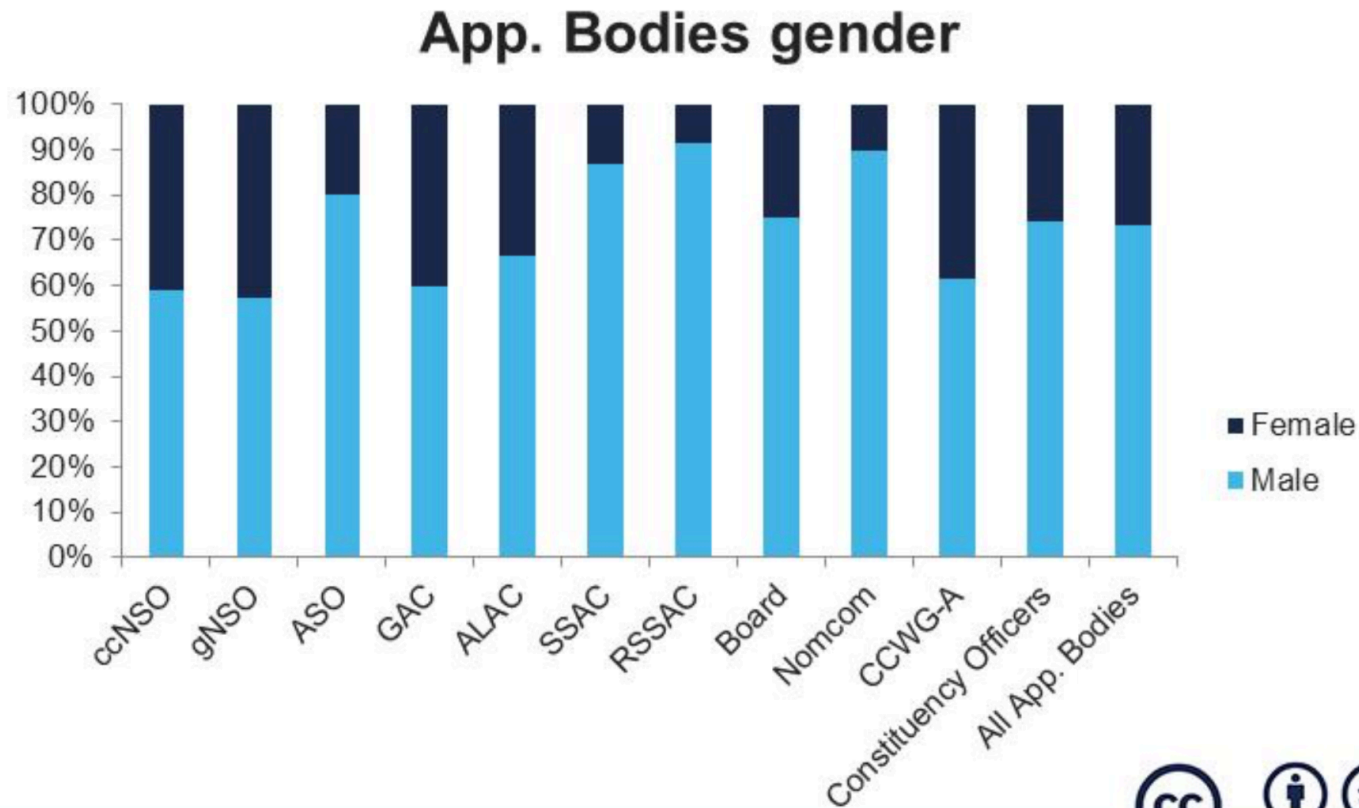
ICANN Diversity Data

Providing facts and figures to support diversity enhancements

20 June 2016

Example output

4.3.2. Some organizations within ICANN achieve a reasonable level of gender balance





Challenge

- ICANN already has strong commitment to openness
- Needs to expand this to the modern age of data-driven organisations



Cleaner and safer DNS





Very active area

- Highly developed use of data
 - Multiple research teams, cooperative forums, NFP services and commercial providers
 - Multiple data resources, extensive data sharing
 - Tools: Entrada, Turing, hadoop-pcap, zonemaster
- Data collection – source evidence
 - Passive monitoring (e.g. DNSDB, PassiveTotal)
 - Hand produced by threat researchers
- Strong sharing culture – via data feeds
 - 40+ feeds available (both NFP and commercial)
 - Track domains, IPs, credentials, URLs, etc

Case study – threat sharing

- Data shared with cooperative forum – then shared with registrars

NZ websites with [redacted] being shared on [redacted]:

Domain	Vulnerable	IP Address	Net Provider	Registrar
wonderworldpetcentre.co.nz	Yes	119.47.114.165	Umbrellar/Web Drive	Domain Agent
samedayprint.co.nz	Yes	210.55.4.23	Spark	RegisterDirect
footes.net.nz	Yes	103.208.86.25	Zappie Host	Domain Agent
karaoke.co.nz	Yes	72.34.61.88	IHNetworks (US)	Domain Agent
hearwell.co.nz	Yes	202.124.98.181	2Degrees	Freeparking
laserdirect.co.nz	Yes	72.34.61.88	IHNetworks (US)	Domain Agent
odarid.co.nz	Yes	192.185.117.48	WebsiteWelcome (US)	1st Domains
cleaningproducts.net.nz	Yes	192.185.117.48	WebsiteWelcome (US)	1st Domains
natureplaynz.co.nz	Yes	192.185.147.101	WebsiteWelcome (US)	1st Domains
napiernz.com	Unsure	103.14.218.170	theCloud.net.nz	Melbourne
theirisboutique.co.nz	Offline	210.79.48.7	Umbrellar/Web Drive	Domain Agent
pestrid.co.nz	Offline	192.185.117.48	WebsiteWelcome (US)	1st Domains
steeleroberts.co.nz	Offline	109.73.228.143	SiteGround (SG)	1st Domains
4myearth.co.nz	Offline	103.6.29.2	hd.net.nz	Freeparking



Challenge

- Those of us “in the know” are “in the know”
- How do others get access to threat data and use that for their organisation?
- Do our trust models scale as the industry grows?
- Is any central coordination and/or cataloguing needed?



Business - more, new, better





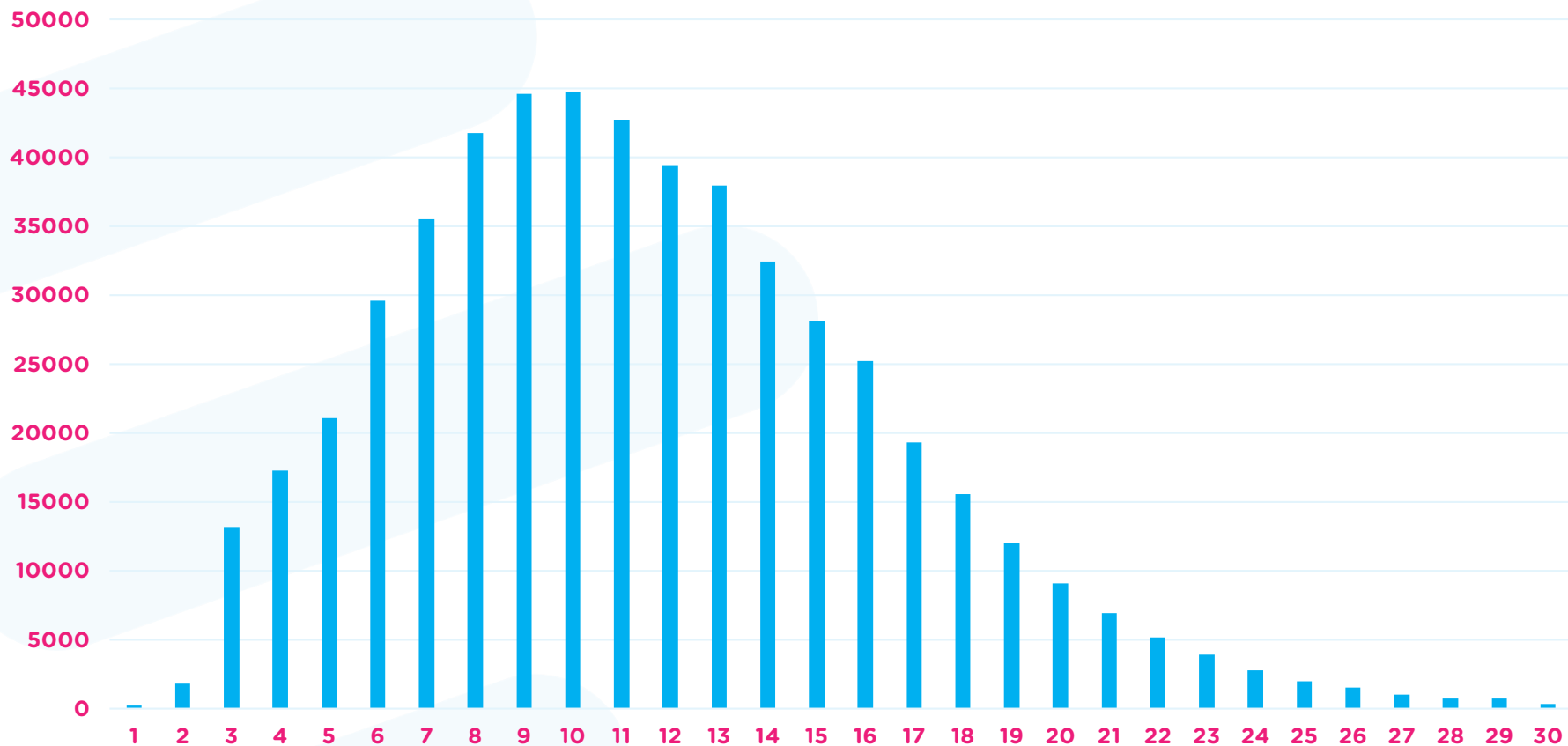
Data driven business

- Hard truths of the domain name market
 - Slowing market – growth is hard to find
 - Strong competition
 - Significant deficit of innovation over many years
 - Danger of registrants thinking domains are ‘stale’
- We need to raise our game to combat this
 - Better market intelligence
 - Targeting marketing
 - New products, same customers
 - New products, new customers

Market intelligence - basic

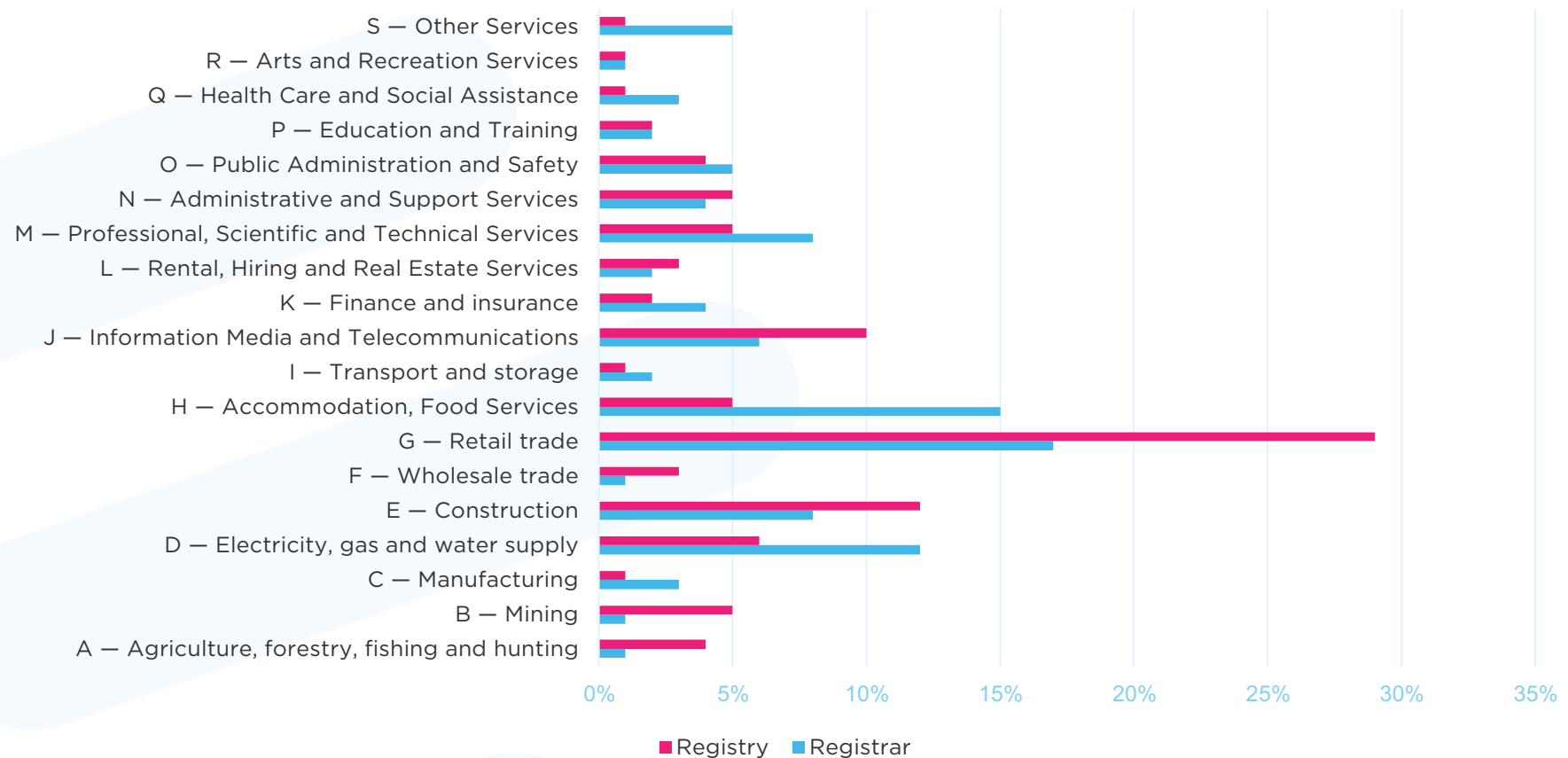
- "Registrants prefer shorter names" - right?

Number of domain names by number of characters



Market intelligence - advanced

- Domain name categorisation by industry



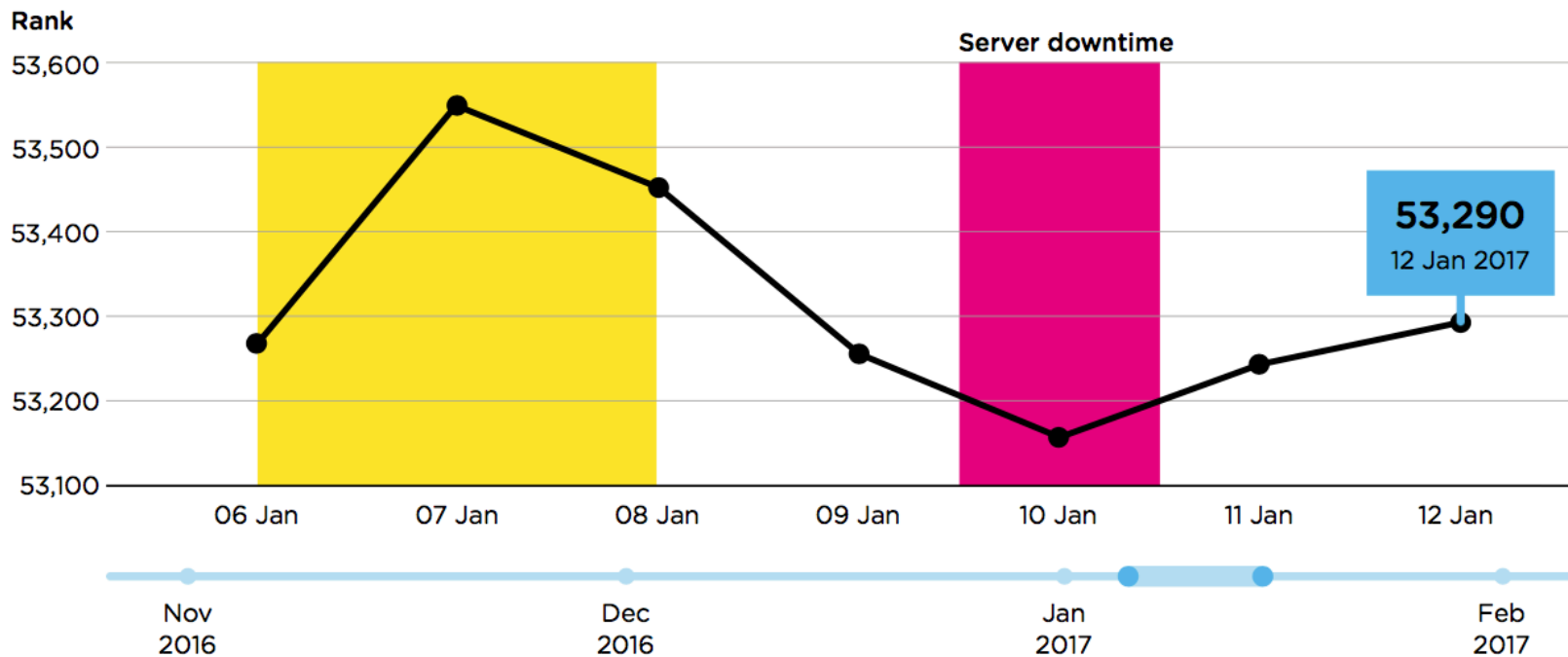
Targeted marketing

- Domains most likely to renew for 10 years
Those in top 20% by observed traffic
- Domains in danger of cancelling
No MX record
- TLD cross-sell opportunities
Simple data matching
- Industry verticals
Machine learning classifier using web site text
- Expiring domains
Valued by algorithm

New products, same customers

Domain traffic score

Using the billions of queries recorded by the authoritative .nz servers and a sophisticated algorithm, we extract the underlying domain traffic score for a domain. This is then calculated for each of the services you define and shown in the chart below. Depending on the type of service selected, this traffic score should correlate with other measures such as visits to your web site or emails received.



New products, new customers

- SaaS product market sizing
Specific DNS record indicators for each product
Counted by regular zone scans
Data sold for competitor analysis

Name/Host/Alias	Time to Live (TTL*)	Record Type	Priority	Value/Answer/Destination
Blank or @	3600	MX	1	ASPMX.L.GOOGLE.COM
Blank or @	3600	MX	5	ALT1.ASPMX.L.GOOGLE.COM
Blank or @	3600	MX	5	ALT2.ASPMX.L.GOOGLE.COM
Blank or @	3600	MX	10	ALT3.ASPMX.L.GOOGLE.COM
Blank or @	3600	MX	10	ALT4.ASPMX.L.GOOGLE.COM



Challenge

- Must innovate or things will get worse
- Innovation is a major cultural change
- Balance between competition and cooperation
CENTRstats is a great example of cooperation
- Build an industry around this
Dataprovider already have a booth
- Adopt common standards
Put the registrant experience first



Societal impact



Telling our story

- Things we just don't tell the world:
- How many people employed in DNS
- How much business it generates
- How many communities are empowered
- How much diversity is supported
- The global engagement we enable
- The charitable projects many of us support

Case study – Web Index



74%

net neutrality



62%

empowerment



74%

gender



84% of countries

do not have effective laws and practices to protect the privacy of online communication

Case study - ISOC

Publications

Annual Review

IETF Journal

ISP Column

Member Briefings

Discussion Paper Series

Global Internet Report

Internet Measurement Project

Document Library

[Home](#) » [Publications](#) » Internet Measurement Project

Internet Measurement Project

The Internet Society is pleased to introduce the Internet Measurement Project, in order to determine the impact of technical, policy, and development efforts on the Internet, and the broader impact of the Internet on the economy. This Project provides a unique set of links to a variety of our own and third-party data sources and reports.

We are making this page available for three purposes.

First, as a public service to learn more about the Internet, in hope that everyone finds a new interesting data source or report.

Second, in order to help identify gaps in data, and determine how best to fill those gaps.

Third, to promote new analysis and insights by everyone in the Internet community, including ourselves, to further stimulate a better understanding of the Internet and its Economy.

- 204 datasets only one from ICANN (NRO)!



Challenge

- Not much of a challenge
 - Gather the data
 - Write the report
 - Tell our story
 - Tell our story
 - Tell our story
- Just imagine
 - ICANN Global Industry Report
 - DNS Contribution to Society Index



Putting this into practice





Steps to make it happen

- Commit – employ data specialist
- Begin cultural change
 - Broaden the principle of openness to include data
 - Set the vision of the benefits
- Engage community – “social license”
 - Community expectations of openness vs privacy
- Put data governance framework in place
 - Adjust contracts, policies and processes to support open data
 - Determine privacy protection rules



ICANN's Open Data Initiative

Edward Lewis | ICANN 58 | 16 March 2017

ICANN
COMMUNITY FORUM

58

COPENHAGEN
11-16 March 2017



Agenda

- ⦿ What is ICANN's “Open Data Initiative”?
- ⦿ What are the goals of the initiative?
- ⦿ How does the initiative fit in other efforts?
- ⦿ What are the components of the initiative?

What is ICANN's Open Data Initiative?

- ⦿ An effort to bring “Open Data” to ICANN
- ⦿ ICANN generates and collects data, e.g.,
 - Generated: Monitoring Service Level Agreements
 - Collected: Monthly reporting by contracted parties
- ⦿ Most of this data should probably be public

What is Open Data?

- ⦿ Taken from <http://opendatacharter.net>
 - ⦿ Open by default
 - ⦿ Timely and comprehensive
 - ⦿ Accessible and usable
 - ⦿ Comparable and Interoperable
 - ⦿ For Improved *Stakeholder*¹ Engagement
 - ⦿ For Inclusive Development and Innovation

1. "Stakeholder" replaces the original text, written for government settings

What are the origins of the Initiative?

- ⦿ ICANN community has, over time, requested open access to ICANN managed data sets
- ⦿ There are precedents in the domain name industry, e.g., some country code TLDs provide open data access
- ⦿ Open Data has been on internal project wish lists for some time

What is the challenge for the initiative?

- ◎ ICANN has collected different kinds of data in different formats for different reasons
- ◎ No Document Management Plan has been in place
 - Distributed curation
 - Disparate formats
- ◎ Open Data Initiative must be prioritized with all other projects, including those to fix data-related technical debt

What are the goals of the Initiative?

- ⊙ Ultimate goal: *where possible*, provide access to all data sets
 - ⊙ Limitations:
 - ⊙ Privacy, personally identifying information, policy and contractual obligations, etc.
 - ⊙ Re-publishing data acquired from third parties with constraints
- ⊙ Near-term: get to the ultimate goal within resource realities
 - ⊙ Pilot programs
 - ⊙ Design process, select appropriate tools
 - ⊙ Prioritized sequence

How does the Initiative fit with ICANN activities?

- ◎ The Open Data Initiative is not operating in a vacuum
 - ◎ WHOIS Accuracy Reporting System is considering open data in its work
 - ◎ Projects to bring in document management processes
 - ◎ Collaboration with custodians of data across the organization

How does the Initiative fit with ICANN activities?

- ⦿ The Open Data Initiative is intended to support community work
 - ⦿ Be responsive to community requests
 - ⦿ Anticipate community requirements
- ⦿ Increase openness and transparency

What are the components of the Initiative?

- ⦿ Identify ICANN managed data
 - ⦿ What limits exist on openness?
- ⦿ Deploy a pilot
- ⦿ Determine process for making data public
 - ⦿ Look for appropriate tools (commercial or open source)
- ⦿ Listen to the community for prioritization hints
 - ⦿ Direction on how the data is delivered

Engage with ICANN



Thank You and Questions

Reach us at:

Email: engagement@icann.org

Website: icann.org



twitter.com/icann



facebook.com/icannorg



youtube.com/user/icannnews



linkedin.com/company/icann



soundcloud.com/icann



weibo.com/ICANNorg



flickr.com/photos/icann



SlideShare

slideshare.net/icannpresentations