

Internet Library of Assigned Names and Numbers

Disorganised Handwaving

Joe Abley, San Juan, March 2024

What data?

ICANN's Data Exhaust

- ICANN and PTI maintain and publish various datasets relating to Names and Numbers
 - Number Assignments and related metadata
 - Contents of the Root Zone and related metadata
 - Contents of various TLD zones as contributed to CZDS
 - Operational metrics, e.g. hedgehog/dns-stats.org, IHTI

When data?

- Many/most of these datasets are presented as a point in time (time of access)
- The history of these datasets is collated by external researchers in some cases, often with duplication of effort, sometimes due to the terms of data access agreements
 - CZDS history has been used in various research projects
 - Root zone history is recorded by many independent archives
 - Some operational data is aggregated over time, so raw data doesn't persist but changes over time are visible over some axes

Why data?

Who cares?

- It is easier to learn from history if the history you care about has been reliably recorded
 - It is difficult to predict precisely what history you're interested in until you realise you have a question
- There is an advantage to having a well-curated set of data available
 - Well-curated relates to data governance, and concepts like authenticity, provenance
- It doesn't seem like that much of a stretch to find security and stability arguments for data archival and preservation

Who data?

Whose library?

- Datasets that are worth preserving and which are replicable without sacrificing data governance goals can be archived in multiple places
 - This is perhaps a goal; perhaps being able to measure the degree of replication as a measure of archival stability is important
 - Data governance still starts at collection
- Perhaps ICANN has a role here, to make data available in a way that allows appropriate governance that encourages preservation
 - What is the point at which data can be considered part of the historical record?

Internet Archaeology

Preserving the wider fossil record

- Jim Cowie (Renesys, Dyn, DeepMacro) has been thinking about this
 - https://content.cooperate.com/post/internet_history/
- Digital Library Science is a thing
- Cryptographic attestation for data integrity and subsequent transformations are finding applications in other areas
- Many other associated datasets
 - e.g. RIPE RIS, the Route Views project, DNS-OARC's DITL

Next Steps

Can and should SSAC look at this?

- It feels like there is some useful work to be done, if we assume that a historical record is worth having
 - Identification of ICANN-related datasets
 - Convening a conversation around historical record vs. operational or commercial sensitivity
 - Framing a broader investigation into initial publication and the choices that would constrain opportunities for useful preservation
 - Coordination and socialisation with other parts of the community and org