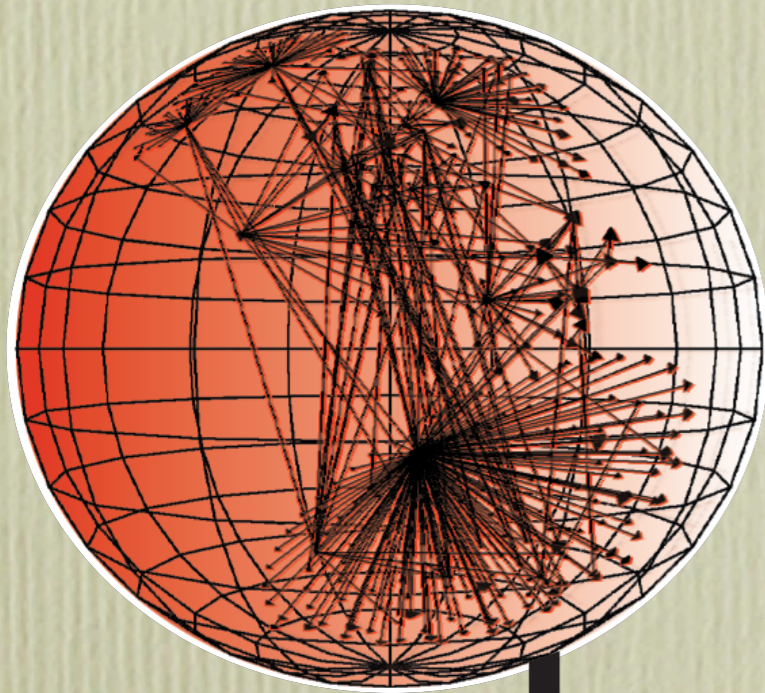


# DITL2008 and DatCat

---

Emile Aben <emile@caida.org>  
10th CAIDA-WIDE workshop  
Aug 15-16 2008, Marina Del Rey, CA, US



caida

www.caida.org

# DITL2008

- Day in the Life of the Internet (DITL)
  - 48hr Internet Measurement event
  - March 18-19 2008 (UTC)
  - Summary at: <http://www.caida.org/projects/ditl/summary-2008-03/>
    - 8 of 13 DNS roots participating
    - Compared to 2007:
      - Double the number of participants
      - Double the volume of DNS data

# DatCat

- Internet Measurement Data Catalog ( DatCat ): <http://imdc.datcat.org>
  - Facilitate searching for and sharing of data among researchers
  - Enhance documentation of datasets via a public annotation system
  - Advance network science by promoting reproducible research
- No data storage, only meta-data:
  - How many IPv6 pkts in this pcap file?
  - How was this pcap file created?
  - How to get access to data

# Lesson learned from DITL2007 and DatCat

- Researchers need data described in great detail:
  - Find out if data is useful for them or not
  - Find the most useful data
  - Can do research on meta-data
- Data contributors usually don't have time to describe data in detail

# (Temporary) Solution

- Add a curator-role:
  - Fills out general information about the data
  - Extracts metadata when data is accessible (e.g. OARC)
- Data contributor checks the general information, corrects it, fills in the blanks, adds extra info
- Curator processes this into a submission

# Result

- Less work for contributor
  - Only needs to check
- Extra role: curator
  - Create initial descriptions
  - Extract metadata
- Takes time/coordination

# Interface for curator and contributor to collaborate

DatCat: Edit DITL Submission: /contact/1-060N-1=Sebastian+E.+Castro ditl2008: collection


In DatCat, data is organized in collections. Please fill out the form below with information about this particular collection. Please replace strings delimited by double asterisks with your input as appropriate, and please add any additional information you deem relevant.

For a full example of an existing catalog entry, see [K-root DNS traces for DITL 2007](#).

Fields marked with \* are required.

|                              |  |          |
|------------------------------|--|----------|
| *Name                        | NIC Chile, .CL ccTLD DNS trace DITL 2008   | 40/128   |
| Keywords                     | DITL, DITL-2008-03-18, ccTLD, OARC, DNS, passive, trace<br>Some suggestions: DITL, DITL 2008,<br>For a full list of keywords already in use, see <a href="#">Browse Collections by Keyword</a> .   | 55/500   |
| *Short description           | DNS PCAP traces containing DNS queries, collected at 4 .CL nameservers.  | 71/128   |
| *Summary                     | DNS packet traces collected during the DITL 2008 event. Includes queries from 4 authoritative nameservers for .CL ccTLD. The trace files in this collection contain IPv4 traffic with DNS queries (no replies), and a total of 75.27 million packets.<br><br>This is a description that summarizes in a couple sentences the most important features of this collection, this can include some summary statistics.   | 246/800  |
| Motivation                   | This data was collected as part of the DITL project, in which DNS and other Internet data is collected for a 2-day period. As a member of OARC, NIC Chile participated for the first time.   | 186/400  |
| *Description                 | The traces in this collection were collected at four authoritative nameservers for .CL located in Chile. Three of them are members of the same anycast cloud (tucape, santiago, valparaiso), the other is a unicast node (ns.nic.cl). The unicast node serves other Latin American ccTLDs, but this traffic was not captured. The collection procedures were thoroughly tested on the days previous to the collection event, so no problems of disc space, clock skew were present. Only an additional<br><br>Comprehensive description of this collection. Some suggestions of information to include here for ccTLD DNS pcap data: <ul style="list-style-type: none"> <li>Information on your setup (number of nodes, anycast/unicast etc.), if you have a URL describing the setup please provide it</li> <li>Information about problems and/or human errors during collection of the data</li> </ul> | 657/4000 |
| Description URL              |  | 0/1018   |
| Additional notes to reviewer | If you have a web page describing this dataset, please provide the URL here, otherwise leave this blank.   | 0/4000   |

Software version 1.7.24  
Page generated at 2008-08-05 08:07:34 PDT  
Request processed in 0.085 seconds

 cooperative association for internet data analysis  
based at the University of California's San Diego Supercomputer Center

# Result in DatCat

DatCat: Collection: NIC Chile, .CL ccTLD DNS trace DITL 2008

**Collection: NIC Chile, .CL ccTLD DNS trace DITL 2008**  
DNS PCAP traces containing DNS queries, collected at 4 .CL nameservers.

Jump to: [Description](#) | [Annotations](#) | [Citation](#) | [Record Details](#)

## Collection Contents

- data objects: [192 directly contained](#), [192 total](#)

## Collection Details

**Summary** DNS packet traces collected during the DITL 2008 event. Includes queries from 4 authoritative nameservers for .CL ccTLD. The trace files in this collection contain IPv4 traffic with DNS queries (no replies), and a total of 75.27 million packets.

**Motivation** This data was collected as part of the DITL project, in which DNS and other Internet data is collected for a 2-day period. As a member of OARC, NIC Chile participated for the first time.

**Data Start Time** 2008-03-17 17:00 PDT (-0700)

**Data End Time** 2008-03-19 16:59:59.999 PDT (-0700)

**Data Duration** 1 days 23:59:59.999 (172 799.999 s)

**Creators** [Mauricio Vergara](#), [Sebastian E. Castro](#), [Duane Wessels](#) (1)

**Primary contact** [NIC Chile DNS Admin](#)

**Keywords** [ccTLD](#), [DITL](#), [DITL-2008-03-18](#), [DNS](#), [OARC](#), [passive](#), [trace](#)

**Used in publications** (none)

**Member of collections** directly contained by

- [Day in the Life of the Internet, March 18-19, 2008 \(DITL-2008-03-18\)](#)

also contained by

- [Day in the Life of the Internet \(DITL\)](#)

**Description** The traces in this collection were collected at four authoritative nameservers for .CL located in Chile. Three of them are members of the same anycast cloud (tucapel, santiago, valparaiso), the other is a unicast node (ns.nic.cl). The unicast node serves other Latin American ccTLDs, but this traffic was not captured. The collection procedures were thoroughly tested on the days previous to the collection event, so no problems of disc space, clock skew were present. Only an additional step was added to the process: all traces, once collected, were sent to a transfer server for upload, due to bandwidth constraints on the nodes where the collection ran.

## Annotations

## Citation

Please use the following BibTeX citation to cite this collection. Some parts are **optional** or may need to be **edited**. To use the “`\url{...}`” command for nice URL formatting, you must call “`\usepackage{url}`” in the LaTeX preamble.



# Current status

- Positive reactions from contributors to reduced workload
- Contributors replace strings of text when asked to, but generally don't add to text
- 17 datasets indexed in DatCat
- 11 datasets currently in process
- At least 7 OARC datasets and 3 others to be done

# Longer Term Solution

- Have ability for DatCat contributor to describe a collection of data, but not necessarily all files in it
- Contributor can add detailed description of data later (at the file level)
- ... but probably needs incentives to do so

# Links

DITL <http://www.caida.org/projects/ditl/>

DatCat <http://imdc.datcat.org/>

DITL2008  
summary <http://www.caida.org/projects/ditl/summary-2008-03/>

DITL2008 in  
DatCat <http://imdc.datcat.org/collection/I-05MM-F>