



# DatCat

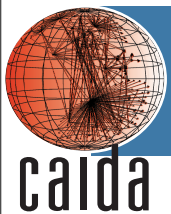
## Overview/Lessons Learned

cooperative association for internet data analysis

Bradley Huffaker <[bradley@caida.org](mailto:bradley@caida.org)>

GEC7 - 17 March 2010





# Goal



DatCat was designed to improve data sharing by providing a unified metadata database for Internet data.


<http://imdc.datcat.org>



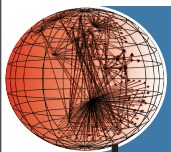
# Goals



## Makes the following processes easy for users.

- finding data sets of interest
  - Many researchers lack access and/or expertise to collect data needed for their research.
- adding new data sets to the catalog 
  - Contributors (who are generally underfunded and providing data out of dedication to the general good) want to minimize time lost to their own research.
- annotating data sets in the catalog
  - Provides a flexible way for contributors and users to mark up interesting facts about data sets. Such as the number of packets or that a given file is corrupted.
- does not store data

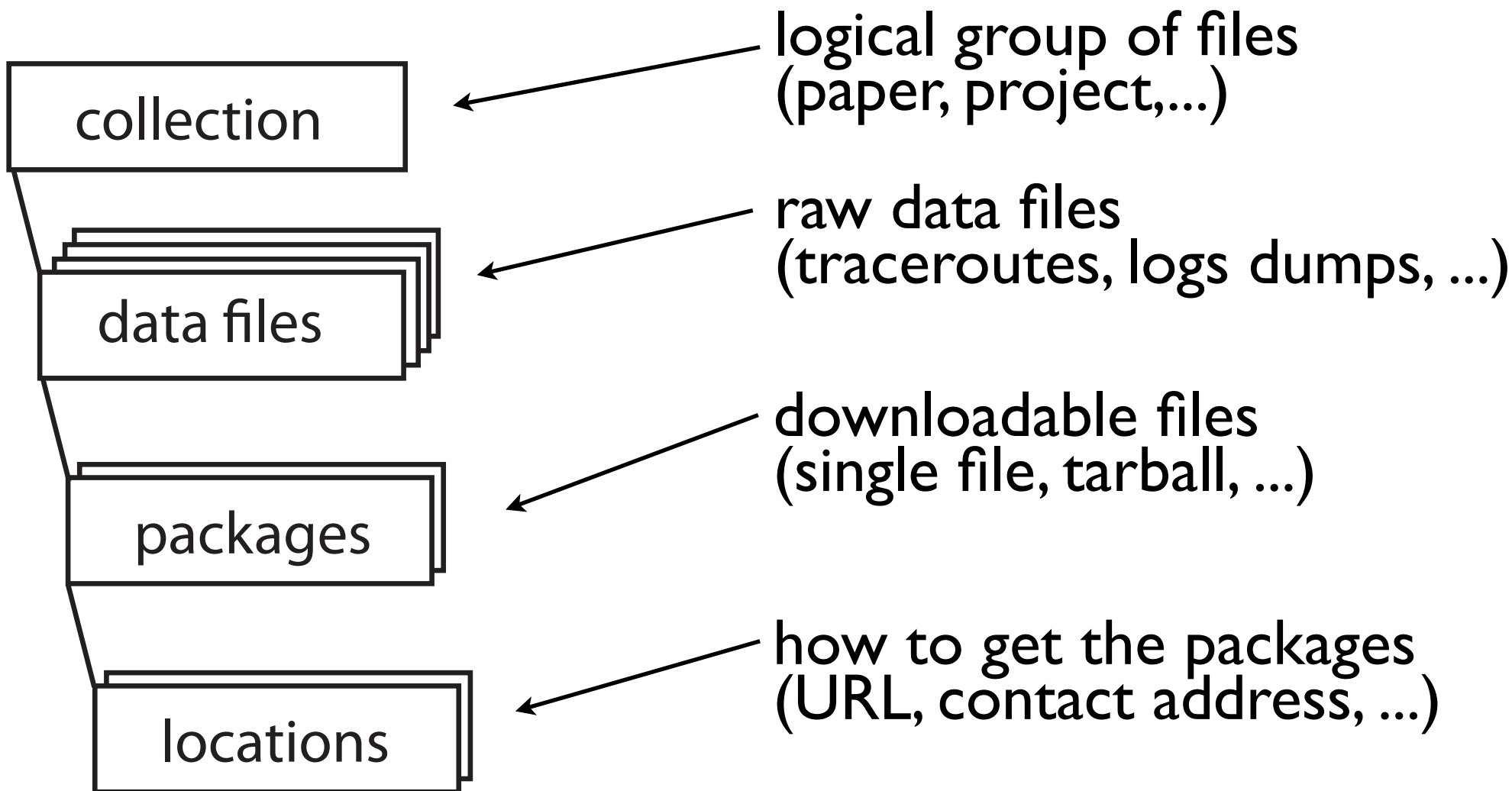
<http://imdc.datcat.org>



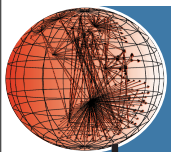
# Database Scheme



caida

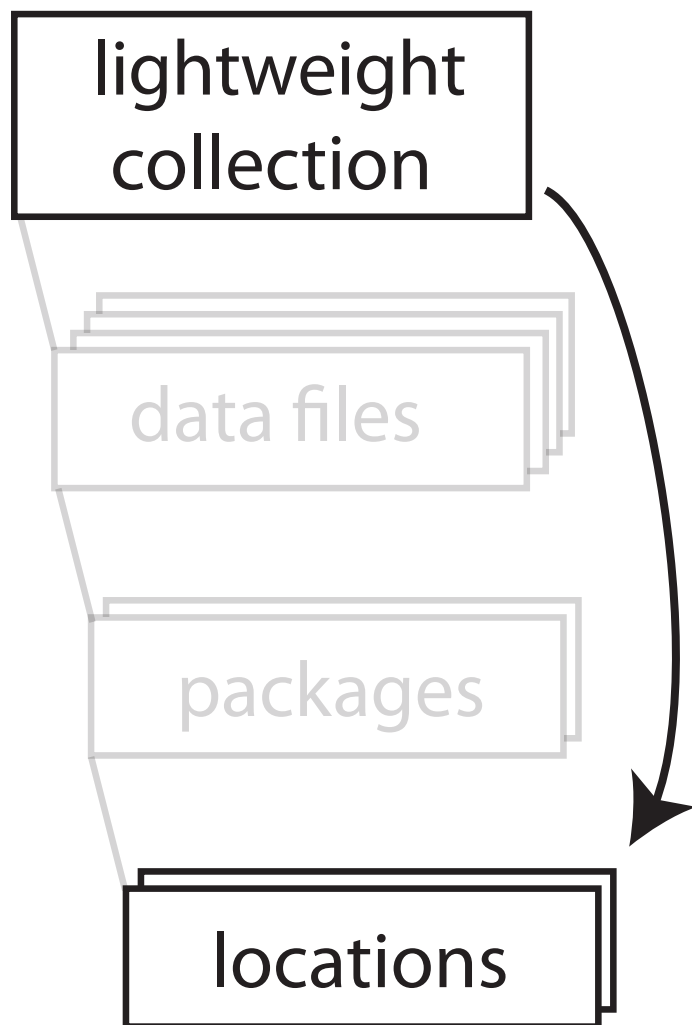


<http://imdc.datcat.org>



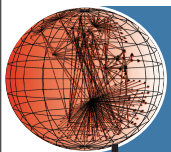
calda

# Lightweight Collections



- contributors will only need to fill in the collection information, plus
  - creation process
  - location
  - data files optional
- shorten search path from collection to locations

<http://imdc.datcat.org>



# Example



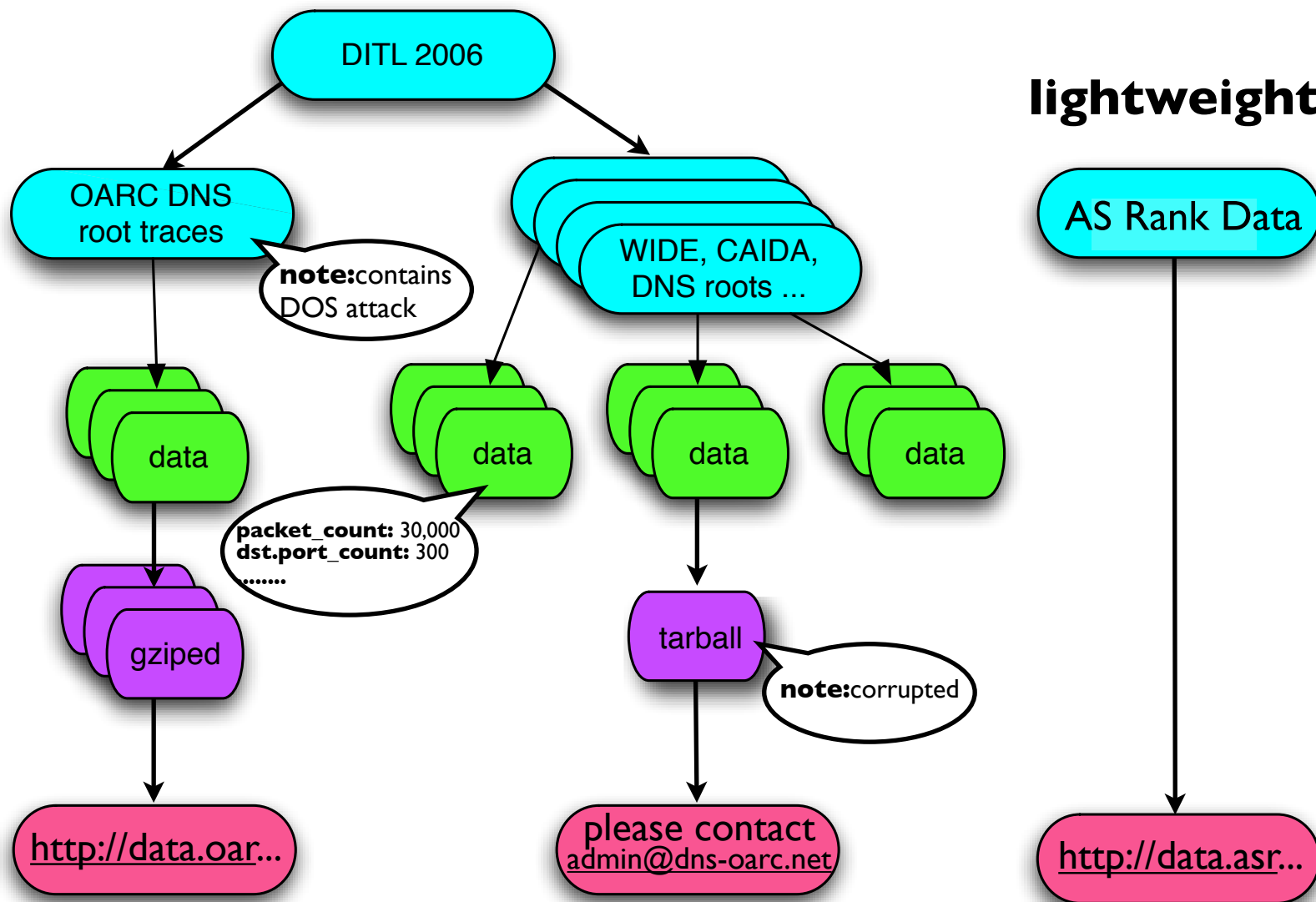
caida

collections

data files

packages

locations



lightweight

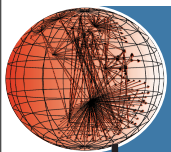
AS Rank Data

<http://data.oar...>

please contact  
[admin@dns-oarc.net](mailto:admin@dns-oarc.net)

<http://data.asr...>

<http://imdc.datcat.org>

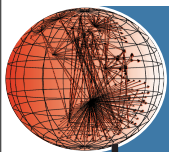


# Collection Fields



caida

* Collection name ?	Macroscopic Internet Topology Data Kit (ITDK) 2012-07	53/128
* Short description ?	The ITDK contains data about connectivity and routing gathered from a large cross-section of the Internet.	95/128
Creators ?	<input type="text" value="Brad Huffaker"/> <input type="text" value="Ken Keys"/> <input type="text" value="Young Hyun"/> <input type="button" value="add another"/>	
Primary contact ?	<input type="text" value="handle or partial name"/>	
Keywords ?	topology, routing, alias resolution	60/500
* Summary ?	<p>The two included router-level topologies are generated from the same IP-level topology but differ in the accuracy and completeness of the alias resolution performed to create them. The first topology is derived from aliases resolved with MIDAR and iffinder, which yield the highest confidence aliases with very low false positives. The second topology also uses MIDAR and iffinder but further includes aliases</p>	183/800
Motivation ?	<p>This dataset is useful for studying the topology of the Internet at the router-level, among other uses.</p>	103/400
Data start time ?	2012 -07 -dd hh:mm:ss.mil -08 US/Pacific	
Data duration ?	<input checked="" type="radio"/> <input type="text" value="5"/> days <input type="radio"/> ongoing	



caida

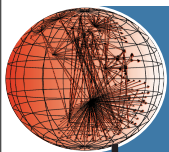
# Location Fields



Fields marked with \* are required.

*	For collection	<a href="#">Macroscopic Internet Topology Data Kit (ITDK) 2012-07</a>
	Creators ?	<input type="text" value="handle or partial name"/> <input type="button" value="add another"/>
	Primary contact ?	<input type="text" value="handle or partial name"/>
*	Status ?	<input type="text" value="active"/>
	Download procedure ?	<pre>&lt;a href="http://www.caida.org/data/active/topology_request.xml"&gt;Request Access&lt;/a&gt; to the <u>CAIDA</u> Internet Topology Data Kits and other Topology Datasets</pre> <p>0/1018</p> <p>Markup: <input type="radio"/> none <input checked="" type="radio"/> HTML</p>
	Download URL ?	<input type="text" value="http://www.caida.org/data/active/topology_request.xml"/> 0/1018
	Geographic location ?	<input type="text" value="San Diego, CA, US"/> 0/4000
	Logistic location ?	<input type="text"/> 0/4000
*	Availability ?	<input type="text" value="restricted"/>
	Private ID ?	<input type="text"/> 0/256





caida

# Give it a try



## Internet Measurement Data Catalog

[Home](#) | [Browse](#) | [Search](#) | [Submit](#) | [Help](#) |

You are not logged in. | [Log in](#) | [Create an Account](#)

[Contact us](#)

### Welcome to DatCat<sup>SM</sup>

The **DatCat** catalog indexes Internet measurement data.

**DatCat** lets you **find**, **annotate**, and **cite** data.

The goals of the system are:

- to facilitate searching for and sharing of data among researchers
- to enhance documentation of datasets via a public annotation system
- to advance network science by promoting reproducible research

For more information, see the [general documentation](#).

#### News

2013-02-05 The new DatCat [simplified online submission system](#) is now available.

2013-02-15 DatCat is back! All previously existing information is intact and available for browsing. We are in the process of reorganizing the structure of DatCat to make it easier to find and contribute information.

2010-09-17 DatCat database shut down

start here



<http://imdc.datcat.org>