

Zone Transfer Policy Review Submission

From: Ewen McNeill
Received: 6 October 2003

With regard to the current Zone Transfer Policy at:

http://dnc.org.nz/content/zone_transfer.pdf

and the call for comments at:

http://dnc.org.nz/content/zone_transfer_policy_review.pdf

I would like to make the following comments in response to the questions raised. (And as a more general comment I'd suggest that the use of PDF documents for what is almost entirely plain text complicates the process of preparing a response, including eg quoting the questions raised -- any inaccuracies in the questions are due to having to retype them from scratch.)

1. Should any organisation outside of the registry company be permitted to obtain a copy of the .nz zone file?

Yes. Any individual or organisation with a legitimate research purpose should be permitted to obtain a copy of the .nz zone file. The DNC office (perhaps with appeal to some suitably body such as the NZOC) should determine what are legitimate research purposes, with the expectation that no reasonable request will be refused.

Aside from research purposes the only reason for accessing the data in the zone files is operational purposes which are adequately served by individual queries through the "100% reliable" network of DNS servers.

2. If organisations are permitted to obtain a copy of the .nz zone, what criteria should be in place to assess whether their use of the zone file is appropriate?

Applicants should have a (brief) research plan listing the research objectives that they wish to achieve with access to the data, and a brief explanation of why access to the zone file is necessary (or significantly useful) in achieving those research objectives.

They should also indicate any republication of the data they plan to do (either in summary or in aggregate).

By "brief" I mean "20 lines or 1 page would be plenty"; it simply needs to be sufficient to indicate what they hope to achieve.

Where the DNC office is of the view that an objective can be achieved without access to the zone data it may be sufficient for the DNC office to demonstrate how this may be done without the zone data (eg, by reference to statistics published by the

DNC/NZRS); but I do not think that requests for access to do legitimate research should be denied.

3. Should authorised registrars automatically be able to achieve a copy of the .nz zone data?

No. The registrar interface to the registry provides for most of their operational and research needs. If they have other research needs they can apply like any other organisation for access.

4. How frequently should copies of the .nz zone file be available to any approved organisation?

As often as is required to support their research objectives, but no more often. Where frequent access is required it may be useful to try to limit the period over which the access is available (eg, there may be a legitimate research need to have every copy of the zone file over a one week period -- eg, to study registration trends on a new 2LD becoming available -- but the access could be limited to that one week). If the research is to track long term trends then access once a week or once a month may be more than sufficient.

5. What information, if any, should any party receiving the .nz zone file be able to make public?

Any data required to support their research/findings. Because of its publication in the DNS none of the individual information is inherently private, so the only concern is its aggregation.

The release of detailed data (as opposed to statistical summaries) should be limited to the amount required to support the research/findings, in a form that discourages "automatic collection" of the details, and subject to an agreed privacy/data use announcement being displayed with the data (eg as is done with the whois results).

6. Does the release of the .nz zone file have a negative impact on the security of the information held?

No. All the information is individually available already. The only concern is the aggregation and ease of access to large quantities of data as part of a larger "data mining" exercise. This is something that would be considered as part of the research proposal in 2 above.

-- Other general comments

Some responses have suggested reducing the data that is released to just the domain names; that would have the effect of reducing the research uses of the data (or requiring that information to be recovered in another manner -- such as by mass DNS queries), and aiding those who simply wanted a list of domain names to drive other data mining for illegitimate purposes. It is also more difficult to provide. As such I do not see any reason to provide only the domain names.

Some responses have also expressed concern at indicating, eg, that 10 name servers are responding for 58% of all names. I do not see the concern in exposing this information, and see it as a legitimate research output. (If someone wished to target a particular nameserver to take a particular domain "off the air" they could easily establish which ones to target through a normal DNS query anyway; if they wanted to know which ones to target to take as many domains as possible "off the air" they could make a fairly accurate guess without any access to the data simply based on, eg, which ISPs have the most customers (widely known to the first order of approximation). But, eg, trends in reliance on a certain set of nameservers is a useful research item.)

One of the things which prompted this review was the (apparent) use by spammers of data obtained from Mark Davies's "new domain names" lists and query tools. While the statistical information provided there has been useful over many years, the query tools have been increasingly less needed on the modern Internet with its high reliance on the World Wide Web, and WWW based search tools (Google, et al). As such I see the statistical information as more of a legitimate research objective than the query tools (in terms of the criteria above).

In the absence of any wildcard "whois" searches (and having written the "whois" server currently in use I can say that there definitely isn't any wildcard facility) Mark Davies's query tools may still have a legitimate part to play. However if it is to remain the output may need to be changed to a format which further discourages automatic use of its results. (As but one possibility: the results could be returned as an image, instead of as plain text, a technique used by some online signup pages now to reduce automatic usage. Or perhaps it could simply return fewer results each time, and borrow some of the whois server's "defences" against repeated queries.)

Ewen McNeill