

Statement of the Non-Commercial Stakeholder Group on Open Data Initiative Data Sets and Metadata¹

The Non-Commercial Stakeholder Group (NCSG) welcomes the opportunity to comment on the Open Data Initiative (ODI) data sets and the metadata that ICANN intends to publish along with each data set. This is a welcomed move forward for ICANN's bylaw-mandated principles of accountability and transparency, and we support the ODI's concept of bringing comprehensive access to raw data to the ICANN community in order to enable evidence-based policy development. While the publication of raw data in and of itself is a first step toward transparency, it is only when the right data is published in a means that can be meaningfully understood, that there is real value generated from this exercise. In pursuit of this objective, the NCSG has structured our comments to address (1) the proposed data sets, (2) the required data quality, and (3) the resources that the ICANN community requires in order to interrogate this data.

About the NCSG

The NCSG represents the interests of non-commercial domain name registrants and end-users in the formulation of the Domain Name System (DNS) policy within the Generic Names Supporting Organization (GNSO). The NCSG is proud to have individual and organizational members in over 160 countries, and as a network of academics, Internet end-users, and civil society actors, we represent a broad cross-section of the global Internet community. Since our predecessor's inception in 1999 (the Non-Commercial Domain Name Holders Constituency, NCDNHC), we have facilitated global academic and civil society engagement in support of ICANN's mission, stimulating an informed users and building understanding of relevant DNS policy issues.

Proposed Data Sets

If properly executed, we believe the ODI will enhance the transparency and accountability of ICANN's activities. After thorough review, we believe that the data sets that should be published are those which are relevant to current and future topics in the community, practicable in use, and of good quality. They should allow the ICANN community to identify trends, to take note of pain points, and to monitor the effectiveness of activities and corrective actions that have been taken. However before we delving into the technicalities, we note our strong position on the need for ICANN to respect human rights, including the fundamental right to privacy, particularly as it relates to personal data which may be in ICANN's possession.

¹ <https://www.icann.org/public-comments/odi-datasets-metadata-2018-06-11-en>

In technical terms, NCSG's priorities for publishing datasets on the open data platform have been set taking into consideration: (1) available data, and (2) NCSG data needs. We note that some of the data in the spreadsheet was missing, thus our prioritization comes based off of the limited data available to us. We have also flagged the data that we think is of least priority especially the ones that need to be only partially shared where personal information might be made public.

It is of critical importance to the NCSG that first amongst the published data sets be those that contain information related to, but not limited to, (1) accountability, (2) current policy development topics, (3) transparency of ICANN in terms of expenditure, including staff travel expenses, DIDP requests and Ombudsperson case dealings, as well as (4) certain metrics related to their effect on non-commercial and end-users, and (5) data related to ICANN's respect or lack thereof for human rights. The data sets which the NCSG has prioritized, based on the available information in the provided inventory, can be found in Appendix A.²

Data Quality

The NCSG is a strong advocate of evidence-informed policy making; in order to make our public interest-oriented contributions, we require data that is (1) timely and comprehensive, (2) accessible and usable, and (3) comparable, consistent, and interoperable.

Timely and Comprehensive

It is the position of the NCSG that in order for the ODI to be useful, ICANN must publish data in a timely manner according to a publicly published schedule. Data must be complete and a system of version control must be in place. If there are errors which are later corrected, these corrections must be made obvious.

Accessible and Usable

Data must be published in a format that allows for easy access and interrogation. It must be machine parsable (for instance, published in formats such as, but not limited to JSON and XML), provided via APIs, and it must be released under an open data license that permits free use, re-use, and distribution. In addition, it would be an asset for data to be visualised where useful.

² The Titles which NCSG prioritized - based on the available information in the provided inventory - can be found in this [spreadsheet](#), graded as Necessary or High Priority. Also available on the spreadsheet, is the least prioritized Titles that we think should not be on top of the list as they are not as relevant Least relevant Slightly relevant but not of priority and Prioritized but not high on the scale. 5: Necessary - 4: High Priority - 3: Priority - 2: Slightly relevant - 1: Least relevant

Comparable, Consistent, and Interoperable

Data must be published in a consistent and comparable manner to allow for comparable units to be compared with one another with the passage of time.

Community Resource Requirements

Access to data only advantages those who have access to interrogate and analyse said data. Unlike many commercial stakeholders who participate in ICANN processes in order to achieve business objectives, as a community of volunteers the NCSG pursuing public interest-oriented objectives, the NCSG does not have a long list of data scientists who we can call on in order to parse this information. We therefore would like ICANN to make independent data analysis support and training available to us. We believe this request to be consistent with recommendation 10.5 of the Accountability and Transparency Review 2 report, accepted by the ICANN Board in 2014, which called for ICANN to *“facilitate the equitable participation in applicable ICANN activities, of those ICANN stakeholders who lack the financial support of industry players.”*³

Errors in the Data Set Inventory Spreadsheet

In our review of the data set inventory spreadsheet, the NCSG observed a number of issues with the data that we ask be corrected in a future update of this resource. These include:

- the “description field” for some identifiers is missing, for instance in sla-monitoring, transaction-reports, operator-reports, activity-reports, [srt]-expenses, when it is mentioned in the metadata vocabulary document that this field is always required.
- the landing page of the identifier [multiple] is missing, although it is said to be published, and published as a page (for reference: Column A, Row 18 - Appendix A).

There are also **omissions**. Data sets which the NCSG is interested in, but are not mentioned in the inventory, include:

1. Number of emails sent to the ICANN policy support staff off of a mailing list, whether a response was sent to the stakeholder or not, and whether or not ICANN staff took action following this email. This should be sorted by stakeholder affiliation;
2. Full details of any events sponsored by ICANN, including event name, location, date(s), whether the support was one-off or recurring, and the nature of the support (if financial, including figures in USD);
3. Any reporting of inability to participate in an ICANN meeting due to the location of the meeting and/or visas issues;
4. All meeting transcripts including the ones related to closed meetings;

³ <https://www.icann.org/en/system/files/files/final-recommendations-31dec13-en.pdf>

5. Details on ICANN-funded face-to-face meetings, including the number of attendees in each meeting room (physical and online), use of scheduling apps by unique participants, and the date of registration and subsequent badge collection;
6. How many email messages have been sent to the Ombudsperson; how many informal complaints have been received; how many formal complaints have been received; and how many cases have been forward (if any) or rejected; and
7. Tracking effects of policy changes at ICANN.

In terms of the proposed metadata vocabulary, as far as our discussions have gone, we believe the title vocabulary is sufficient enough for a start, however a lot of information is missing under those titles. The fields “keyword” and “theme” are declared as “always required” in the metadata description but keyword field is populated with only ten out of 231 lines, while the column “theme” was never really utilized at all. We would like to better understand the role of this field and the reason to have it if there will be no information crowding it. There are also two titles that are flagged as restricted (tld-zones and L-Root), and we would appreciate a brief explicit reasoning as to why they are restricted.

As this initiative aims to increase the accountability and transparency of ICANN, we are looking forward to the successful roll-out of this initiative. However, there is a need for care and responsibility when personal information is being disclosed (which should not be used as an excuse not to publish information that is in the public interest, such as details on staff travel expenditure), and there is also a strong need for periodic reviews of the ODI, including of its data sets, in case there is necessary information not found in the inventory.

The NCSG is grateful for the opportunity to comment on this issue, and we trust that our comments will be taken into consideration. We are available to clarify our recommendations, if necessary.