



Next Generation of Whois: An Overview

Introduction

The purpose of this paper is to provide an overview of issues raised by the DNS community during the WHOIS and Expert Working Group (EWG) discussions and webinars held during the past many months leading up to EWG releasing their final paper to the ICANN Board.

The next step is for ICANN's policy-making bodies to take up the work and craft a new domain name registration data validation and retention policy. This paper is a guide to the issues raised and will help DNA Members understand and engage in the discussions.

1. Summary

- The Expert Working Group on next Generation Directory Services (EWG) Proposal,¹ if implemented, is likely to have a significant financial impact on registries, registrars and possibly registrants:
 - Costs will increase;
 - The stated benefits are uncertain,
 - Anticipated cost savings for registrars are likely to not materialize,
 - The costing estimates for the centralized Registration Directory Service (RDS) does not take into account all costs.
- There are legal and competition risks inherent in creating a centralized body housing all the data.
- There is a risk that, with increased regulatory burden, small players will drop out of the market, or that registration rates will suffer because of increased costs.

The EWG Report has been presented to the ICANN Board. The next steps are for the GNSO and other ICANN policy making bodies to consider the report as part of the ICANN bottom-up policy making process.

¹ <https://www.icann.org/en/system/files/files/final-report-06jun14-en.pdf>

2. What does the EWG report say?

The Executive Working Group (EWG) decided the current Whois model should be replaced. The proposal can be divided into:

- Improving data accuracy
- Limiting access to full data
- Data storage

Four key concepts are central to the EWG's recommendations:

- Accuracy of **validated** registration data
- Access to that data by **accredited** users
- Privacy for registrants: certain data elements will be **gated**
- Accountability for those who access data for **permissible purposes**

The new model creates three new entities, validators, accreditors, and a central repository (plus several other new actors). Validators will validate registrant data; accreditors will approve the credentials of those seeking to access the data. Privacy and proxy services will be available.

One could envision that each of the following three areas could be divided into separate Policy Development Processes. Taking them in reverse order for ease of explanation:

Data storage: The EWG decided the registration data should be stored centrally. They considered the present model of data stored at registries and forming regional data repositories (syndicated model). They favored the central database over the syndicated model, as it was felt to promise higher standards, enable simpler entry of data from multiple points to create single registration records, and access to data by authenticated parties.

Access to data: The EWG examined the current uses to which Whois data (Registrant Data) is put and accepted them as twelve different "permissible purposes." In other words, the EWG apparently assumed that all current purposes are legitimate, should be continued and facilitated by the new RDS. That will be a policy discussion for later.

Access to some data would be limited. In the RDS, data would be collected and disclosed for permissible purposes. Some purposes require formal accreditation processes. There is anonymous access to limited Registrant ID information (including the registrant email). After that there is a gated model of access to data driven by who you are and purpose of your data use. An example: if are you accredited to view data for legal reasons, then you can view legal contact, but you cannot see the technical contact.

Data Accuracy: The registrant contacts' information can be validated separately from domain name registration itself. Contacts are verified by local validators – a separate process from the domain name registration. Privacy and proxy services will be available through an accreditation process. In addition to privacy and proxy services, there is Secure Protected Credentials – where individuals under threat can apply for anonymous domain name registration.

The report requires every registrant to have each of these purpose-based contacts (they can be the same person, and will default to displaying registrant data):

- Registrant
- Tech
- Admin
- Abuse
- Legal
- Privacy / Proxy (for those types of registrations only)
- Business (for those types of registrations only)

The purpose-based contacts and the limitations imposed are intended to address data protection laws.

3. Strengths of the EWG report

The EWG report advances the debate on WHOIS in a number of areas, in some cases building on the work of the 2012 WHOIS Review Team². For example:

- Recognition that privacy and proxy services should be included into the accreditation regime, with minimum standards for relay, reveal and contact details.
- Recognition that individuals have legitimate expectations of privacy, and that there are legitimate reasons why entities may choose to hide their details from the public directory (e.g., suppression of religious or ethnic minorities, freedom of expression, etc.).

It provides a thorough analysis of the different purposes for which various stakeholders have had to access registration data. It also explores concepts of accountability for access to registration data, and of the registrant giving permission for processing of personal data for stated purposes.

4. Issues for ensuing policy discussion

The first question the EWG was answering was to define or describe legitimate uses of WHOIS data. This question remains unanswered. Instead, the EWG took all the

² <https://www.icann.org/en/system/files/files/final-report-11may12-en.pdf>

existing uses of WHOIS data, accepted them as legitimate, and then set to developing an implementation model for them.

The first question for the policy discussion should be the same. It is not readily apparent that all uses are legitimate. Should private individuals be granted the same investigative access as law enforcement? If so, what are the criteria a private individual must meet to be granted access to data?

The remaining policy and implementation answers flow from the answer to this question. This primary question, unanswered to date, must now be addressed outside the EWG.

The EWG report spans 166 pages, and contains 180 policy recommendations. The level of detail is intense in areas; numerous new acronyms are introduced. The style makes it difficult to determine basic facts, such as what data will be “outside the gate”.

The policy discussion must be evidence-based when considering the EWG’s assumptions and recommendations. For example, evidence is required to support the causality between “more validation” and reductions in abuses.

The key recommendation, offering registrants privacy with gated data, is somewhat undone by having registrant data appear in mandatory fields that will be public (such as legal contact). Also, the dissenting opinion³ of privacy advocate and EWG member Stephanie Perrin anticipates that most registrants, confused by complexity, will end up with their data published by accident.

Given the concern regarding country-by-country variances in privacy laws, a technical “rules engine” is in the proposal to apply national law, based on the registrant’s jurisdiction. The dissenting opinion by Stephanie Perrin points out this seems to be an unfeasible proposition.

Several assumptions in the EWG report must be tested, including a central tenet that most registrants understand and/or care about the importance of keeping domain name records up to date and accurate. Current evidence suggests otherwise: data quality is acknowledged to be poor, and registrants are the source of most of the current contact data. Early experiences of RAA 2013 indicate that registrants are failing to follow through basic validation steps.

What follow are implementation issues that will affect registries and registrars that must be finalized before an implementation can be considered.

³ <https://www.icann.org/en/system/files/files/perrin-statement-24jun14-en.pdf>

4.1 Financial implications: implementation costs for registries and registrars

4.1.1 Costs will increase

On page 58, the EWG Report notes “*Transition and compliance challenges associated with these new data elements must be considered prior to any RDS implementation.*”

These must include the potentially high cost of implementation for existing registries and registrars.

Validation costs will increase; benefits are unclear

Based on an assumption that more validation equals better data and fewer abuses, the EWG is recommending increased validation obligations that go beyond those in the 2013 RAA. This will increase registrars’ costs, both as validators, and in support costs. In the first 6 months of enforcement of the 2013 RAA data validation requirements, there is evidence that over 800,000 registrations have been suspended, and this has “demonstrably created harm” to internet users⁴. Meanwhile, there has been no evidence presented that the 2013 RAA is providing or the EWG proposal will provide the expected benefits. A rigorous cost/benefit analysis is required.

Costs of refactoring databases

The EWG report “policy” recommendations go to significant technical and operational detail. New mandatory data fields are specified (pages 49-56), as is a contact-based data structure (page 69, and principle 77) and a requirement that each Contact ID must be unique (see page 78, section g). While many in the industry might have such a data structure, others will not. Historically, that has been a choice about how they serve their customers and meet compliance obligations. Implementing a new data structure is non-trivial in terms of cost and time. A policy that stipulates a particular technical solution is bound to have severe cost implications for some, and could stifle innovation or otherwise constrain the marketplace.

Support costs - explaining the new requirements to registrants

Currently, when a registrant wants to update their contact details, they visit their registrar or reseller. In the new proposal, validators, not registrars, will be responsible for modifying contact information (principle 85). If their exiting registrar is not a validator, the registrant will have to go to a separate entity to update their contact details. This is likely to increase support costs for registrars, again with an uncertain and unproven expected benefit.

⁴ Elliott Noss, Board with Registrars Stakeholder Group, London 2014, p18-19:
<http://london50.icann.org/en/schedule/tue-board-rrsg/transcript-board-rrsg-24jun14-en.pdf>

4.1.2 Anticipated savings are uncertain

Dismantling current WHOIS provisioning - wasted investment

The vision is that registries and registrars will no longer be responsible for providing WHOIS access, nor for validating registrant data. Will this save costs? Most likely not, as WHOIS provisioning is integrated into other essential, high availability services that will have to be maintained, e.g., domain availability checking, and domain name registration. The capacity to provide high availability services will continue to be required.

Registrars will be validators

The EWG acknowledges that many registrars will choose to be validators. This reflects reality, as many registrars have recently invested in providing validation to comply with the 2013 RAA. Rather than waste that investment or risk losing customers due to increased complexity, registrars will likely continue with the validation function, rather than pay across an element of the domain registration fee to third party validators. In any event, registrars and resellers have a vested financial interest in being able to contact customers for payment and maintaining the one-on-one customer relationship. Therefore, it is registrars who will likely meet the cost of further enhancements in validation, and of interfacing with the centralized RDS provider.

4.1.3 Cost of implementing the centralized RDS excludes industry costs

According to the EWG report, the cost of setting up the centralized RDS is estimated at €0.02 per domain name. The low cost point gives the impression that switching to the new RDS environment will have little financial impact. This is incorrect.

This cost estimate **excludes** the costs incurred by other actors in the ecosystem such as registries, registrars, set up of validators, accreditors, secure credentials providers and other new actors.

The core volumetric assumptions⁵ and budgetary analysis also include some assumptions that should be tested. For example:

- The number of gTLD registries is predicted to be 2,000 by start 2015, rising to 7,000 by 2020. Latest predictions⁶ suggest that any second round for new gTLDs is unlikely to open prior to 2018. The more the number of gTLDs is inflated the less the cost per domain name of the RDS.

⁵ See slide 18 RDS Costs Analysis Full Report IBM June 6 2014
<https://community.icann.org/download/attachments/45744698/RDS%20Cost%20Analysis%20Summary%20IBM%206June2014.pdf?version=1&modificationDate=1402064227000&api=v2>

⁶ See New gTLD Program Reviews and Assessments, published September 22 2014, which predicts work to start on developing Round 2 to run through to end of 2017.
<http://newgTlds.icann.org/en/program-status/reviews-assessments-draft-work-plan-22sep14-en.pdf>

- Number of domain names, predicted to grow from 151 million (start 2015) to 500 million (end 2020). This equates to more than 230% growth of the gTLD space in 5 years. By comparison, growth of global domain names (including ccTLDs) has grown by just 41% in the past 5 years. New gTLDs have not yet provided stimulated growth to the levels predicted in the costing model. By over-inflating the number of domain names, the estimated cost of the RDS per domain is reduced.
- The IBM model estimates the number of reverse and historic searches as 1% of total queries. Both types of search are computationally expensive, and the model indicates that costs could increase by 35% if the percentage of such queries increased to just 3%. It is unclear why IBM believes that a centralized RDS model would not be as sensitive to changes in volume as the synchronized RDS. More information is needed such as metrics on current volumes of reverse and historic searches and their cost.
- The cost analysis does not include clustering for high availability. While clustering may not increase the *relative* costs of syndicated vs. centralized RDS, it would increase the absolute cost of both. Given the volume of queries across the entire gTLD space, it is highly likely that numerous measures to ensure high availability, resilience against cyberattacks and outages will be required.

In summary, the costs of implementing the EWG recommendations need to be understood as they affect the **entire ecosystem**.

4.2 Legal and competition

4.2.1 The risks of creating a centralized body: an irreversible change

The EWG does not give adequate weighting to the risks of creating a monopoly that has responsibility and control over all gTLD registration data. This creation will change the way data is collected and stored in the DNS for the long term. The EWG has emphasized that the RDS body will operate on a cost-recovery basis, and will employ excellent technical security measures. While the EWG makes many good arguments in favor of centralized database, they must be balanced against the risks as described in a careful technical and business analysis which must be undertaken before the proposals can move forward.

4.2.2 Reverse and historic queries: who will store the data?

European data retention law is currently in a state of flux. Recently, the European Court of Justice voided the Directive mandating retention of data. ICANN has granted numerous data retention waivers to registrars.

This will require a discussion of the inclusion of **reverse queries** and **historic (i.e., “WHOWAS”) queries** as mandatory. Incorrectly, these elements are listed by the EWG as features of the *current* WHOIS (see table on p21 ff, and preamble on p19). Although these searches are currently available as paid for, value-add services, neither are currently part of the mandatory WHOIS, nor are they currently provisioned

on live WHOIS services. Both types of search are computationally expensive, and potentially violative of privacy rights in many jurisdictions. It has been said that both types of query would likely fail tests for necessity or proportionality under European laws, and are also incompatible with destruction requirements in data protection laws. Neither of these features are discussed at length in the EWG report, so many issues are unclear, including enumeration of the perceived benefits, how they outweigh the intrusion on fundamental rights, and which entities would be required to hold and store such historic data. Potentially, registrars could incur liability for breach of data protection and/or applicable privacy laws.

- 4.2.3 **Negative consequences - impact on registrar numbers and registration rates**
Increased levels of cost and regulation requirements are likely to lead to registrars to exit the market. A year on from implementation of the 2013 RAA, 35%⁷ of ICANN accredited registrars have yet to sign up, despite the incentive of access to new gTLDs only being available to 2013 RAA accredited registrars. Higher validation costs and requirements may have the consequence of shrinking the registrar market, reducing choice for consumers. Registrars generally operate on thin margins; any rise in costs and prices is likely to depress both new registration and renewal rates.

5. Conclusions

For the upcoming policy discussion:

The EWG was asked to define legitimate uses of WHOIS data. The question remains unanswered and must now be addressed outside the EWG.

The costs and benefits need to be better understood and articulated. Can improvements to the current system achieve the expected benefits without excessive risks and costs?

Registrars have recently gone through the cost of upgrading processes and systems to comply with the requirements of the RAA 2013. It is not clear that the effort has succeeded in improving data quality, or that the benefit is worth the disruption to legitimate domain name registrations already experienced.

Registrar cost increases should be calculated: the validation investment recently undertaken will become worthless and the proposal mandates new data structures.

The legal and competition issues of creating a single centralized body to control and maintain all gTLD registration data have not been adequately explored by the EWG, nor have the costs or legal risks of mandating historic and reverse queries.

⁷<https://www.icann.org/registrar-reports/accredited-list.html> This is not a precise figure, as some registrars have multiple accreditations, and seemed to have upgraded only some to the 2013 RAA.