**Technical Study Group on Access to Non-Public Registration Data**

**Face-to-face meeting; 14-15 Jan. 2019**

*Use cases*

* Use Case #1: query for a name server by its IP address; not redacted data.
* How do the RIRs handle queries of IP addresses?
* RDAP allows a name server search by IP – is that something we need to consider? Yes.
* Use case #2: What does bulk mean? A lot of queries? Reverse searches?
  + 2a: Security and law enforcement require bulk access to domain records, which might include multiple queries, reduced or no rate limiting and reverse search capabilities (i.e. search domain by an attribute available in RDAP)
  + Use unique identifiers to correlate across domain names.
  + Multiple identifiers may exist but don’t have to get into the details
  + Revised use case #2: Authorized users (e.g., security researchers, law enforcement, etc.) require bulk access to domain records, which might include multiple queries, reduced or no rate limiting, and reverse search capabilities (i.e. search domain by an attribute available in RDAP)
* Use case #3: rephrased as in Google doc
* Towards a Policy Framework for Registration Data Directory Services (Crocker)
  + Graph similar to the data fields identified by EPDP, they have not yet discussed access.
* Use case #6: Merit in exploring the idea of hashed data being lodged with ICANN – would have
* Note why we removed use cases, including redundancies –
* Unicorn: Ram’s statement from 1305-1310
* May be an error condition for an operating model in which a query is sent to an ICANN service that is expecting to receive only accepted conditions – may be normal model of operation for server operators. (or is this a use case?)
* For Rys/RRs, if you’re doing an unauthenticated query, we only give you the public data. Are we expecting ICANN to do the same thing if it receives an unauthenticated query, it would only receive public data?
* Does CP have to operate an RDAP service that services unauthenticated/unauthorized uses?
* Question to include in final report: Is ICANN ready to build and manage a system that moderates access requests, at scale, with SLAs etc. attached to it?
* 2:28 pm (1.5 hours after lunch) Question to Göran: Can CPs have visibility into the requests/the justification for the data requested?
* Who can query what and what gets recorded? Need very clear requirements on that before implementation can begin.
* Principle of centralization: logic should live in one place, where all fields, justifications, authentication, authorization.
* Add to charter assumptions: Data holders assume that ICANN will ensure validity of credentials.
* Question to Goran: Is ICANN the sole party giving access?
* Question to Goran: Is ICANN going to be the proxy for all non-public data queries? Taking into taccount the legal and political implications that involves? Or is a distributed model also a possibility?

*Charter: Key Questions/Issues*

Authentication/Authorization

* Question 1: Synchronous, automated approval or asynchronous approval? If so, that’s not a technical question
  + Comes down to what the policy requirement is.
  + May require two systems – one for synchronous/asynchronous
  + Would require a flow diagram to funnel to answers to question 2 or 3.
* 1st 3 questions in this section are versions of the same questions.
* Question 2: Steve’s model of pro forma requests that a client can apply to a request for information so authorization can happen in real time. That would ensure synchronous authorization.
* Question 3: ICANN would create a ticketing system.
* Answer to first 3 questions: Reply can be authorized in real-time using a pro-forma query. Requests that don’t meet pro forma templates, must be escalated to a human interaction with a support queue.
* Question 4. Potential solutions (reference Scott’s draft on using federated authentication. Also e.g. certificates.) Which is the best fit for the requirements? Need to know more about the requirements.

Data Transport/Storage & Audit

* Question 1: Depends on whether ICANN is the proxy for non-public data requests. What do we log and why? Logging driven by a need to audit access requests. Who’s doing the auditing? Does ICANN need to audit Ry/RR logs? Or only its own? Reconcile the two? Need answers in the form of policy recommendations. ICANN and RDAP server both have to log queries. If logs are used for audit purposes, then we need to come up with a standard format, and a way to ship log data to ICANN for reconciliation. JSON logs may be a good interoperable logging format.
  + **Confirm assumption 3: ICANN is the sole party that authorizes access to non-public registration data requests.** **(Question for Goran)**
* Question 2: If ICANN is the gateway, it’s an implementation question for ICANN. **Raises the question of who has access to logs and who maintains logs. Answer: If there’s a central model, then it could be done on a requestor-by-requestor basis. Counter productive in a decentralized model.**
* **Question 3: Whether access is centralized or not; in order to provide authorized parties with access to logs, having ICANN be the party that aggregates logs and provides access is the only feasible model. Would require CPs to submit their log data to ICANN on some schedule and in a standardized format. How third parties can access that date is an implementation choice for ICANN.**

Access Control Protocol

* Question 1: Delegation of authentication and accreditation is on the table, **delegation of authorization is a question for Goran**.
* Question 2: If ICANN is a proxy ahead of CP proxy servers, then ICANN has to do its own authentication/authorization. Or could use a federated model: Authorizing party would be ICANN and generate a signed credential.
* Question 3: By running its own RDAP service. May be a web interface. Need both RDAP and a web interface that allows third parties to manually input a request. May be automated RDAP interface or non-automated.
  + New use case: registrars need to be able to query data belonging to registrars not under their purview for transfer purposes.
  + New use case: Registries and registrars need to be able to query data for abuse mitigation or other similar legitimate purposes.
* Question 4: If there is data that has to be redacted, how do you show that? Are some parties expecting to receive a reason for why the data is redacted? Already available in RDAP JSON responses, for example, as “Object truncated due to authorization.” Or “Due to lack of authorization.” Design of the asynchronous system must have a prioritization component. How to define priority is a policy question.

Performance Requirements

* Question 1: At a minimum, SLA should be same as what CPs will run as required for RDAP. Requirements should be linked to decisions made by contracted parties. If the service is unavailable either synchronous or asynchronous channel, appropriate communication measures must be put in place.
* Questions 5 and 6: Security requirements have to come from policy. Can say that there must be security requirements and it must be addressed by those bodies debating policy. GDPR article 32 provides a security framework. May use that as a set of guidelines. Recommend a current set of standards that may be evolved with policy.
  + **Action: Benedict/John/Tomofumi to send language on current security/resiliency standards, referencing e.g. ISO standards, etc.** 
    - GDPR Article 32 1. recommends:  
        
      1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:  
      1. the pseudonymisation and encryption of personal data;  
      2. the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;  
      3. the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;  
      4. a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.  
      Reference:  
      https://gdpr-info.eu/art-32-gdpr/ [gdpr-info.eu]

*Other Key Issues for Discussion*

Transparency, assignment of responsibility:

* Which logs will be aggregated? Need requirements. Note that logging is an important requirement and specifications required on what to do with RDAP logs. Set the outcome -- not process. I.e. Logging is required for… xyz.

Error conditions

* Should we specify error conditions? Should we design for these errors? Design flow or give response to how to respond to categories of errors.
* How may errors be made? That enumeration may be passed on to a policy group for them to determine outcomes.
* The system operated by ICANN, ought to have a mechanisms when errors (in info collected from registrants or registrars) appear, there is a process to identify the errors and process to correct the errors. System does not necessarily have to do real-time checks, but what do we do when the system identifies errors
* Possible errors: Registrant submits data incorrectly; registrar mistranscribes data;
* Use Case: ICANN internal processes will require sampling data through system to fulfill WHOIS accuracy requirements and identify problems in the system. (out of scope -- this is for the policy discussion)
* Use case: Registrant requests their own data, tied to a domain name. ICANN can use the system to fulfill data subject requests based upon legal obligations.
  + Action: Identify the use case and flag in paper. (there be dragons!)

Accounting, costs, billing

* If there are requirements for logging, that may be a step toward a solution.
* Could bill by query/credential/etc.
* There be dragons
* Billing model is out of band for TSG

Maintenance and evolution/Governance

* RDAP, protocol may evolve, policies may evolve: All must be accommodated in the spec.
* TSG will conclude its work in Aprill. When done, the paper should specify a process for governance and oversight after implementation. Perhaps sent to GNSO for discussion.

Multi-use requests:

* Covered by use cases.
* Does a party need different authorization based on different purposes?
* Shouldn’t link synchronous with single-purpose, multiple single domain queries, covered by a single authorization.
* Action: Will group consider what attributes (claims) to associate/encoded with tokens? Purpose of query, etc. should be done offline.
* Use cases: Though use cases may be similar, the spec may require different profiles depending on purpose of queries.
* Scott H. walked team through Open ID.
  + How do you take the identity of who’s asking for the information and pass all the way to ICANN (as operator of this service)?
* User journey: User should be able to discover the base URL for the centralized access and authorization system
  + Options for moving forward:
    - Requires another IANA registry? May not be best option.
    - 1. Establish a new link relation type for URL discovery
      * Name for new link relation type: Private
      * Will point to URL of the ICANN RDAP service for non-public data.
      * URL for link should point to specific resource for that domain.
    - 2. Establish a new remark type for URL discovery/description
      * Link to the web page describing the process to be accredited and obtain access to non-public.
    - Recommendation to do both.
* Use cases
  + Use case #1: Reverse lookup is a point of ongoing conversation.
  + Describe operational models when introducing use cases to community.

*Discussion with Goran:*

* Is ICANN going to be the proxy for all non-public data queries?
  + WHOIS is in our bylaws and we have a responsibility for providing that service, so we think we have a responsibility to that.
  + Taking on risk doesn’t mean we get the same risk as another party, i.e. Google.
  + The theory is that CPs’ liability may be diminished if all queries for non-public data go through ICANN servers.
  + Is ICANN willing to publish a transparency report for those logs that would be aggregated in such a model?
    - Law says you have to log. Don’t see a problem with it as ICANN is transparent. That answer has to be a part of any community-developed policy.
* Is ICANN the sole party authorizing access to non-public data?
  + Someone else is authorizing users, i.e. WIPO for IP lawyers.
  + Every request for data has to come with a reason.
  + Open to other solutions to this, including others authorizing requests for data.
  + Ultimately ICANN has to be the only one who transmits the query to the CP. No other party.
  + ICANN has a responsibility for request, as we are sending it.
  + Authorization means that given a requestor and what they want, can we say yes or no to a particular request?
  + Envisaged two systems: \
    - 1. like a web browser, provide domain name, justification and get a token for a limited period of time to access the data.
    - 2. Blanket approval for a specific purpose, but without providing a specific domain name.
  + If you have alternatives, include them in the proposal.
* Can CPs have visibility into the request/justification for the data requested?
  + That’s the transparency report.
  + As long as it doesn’t increase the risk to the contracted parties.

*ACTIONS:*

1. Add to charter assumptions: Data holders assume that ICANN will ensure validity of credentials.
2. Enumerate list of errors that can happen in data collection (registrants and registrars). The outcomes of what should be done with these errors should be referred to policy decision makers
3. Recommend that ICANN have a web portal to report on collection errors. In the RDAP process, provide a response with a link to the process
4. Group should consider what attributes (claims) to associate/encode in tokens? Who/why/what (data fields)/by when?
5. Who has access to logs and who maintains logs. Answer: If there’s a central model, then it could be done on a requestor-by-requestor basis. Counter productive in a decentralized model. (See Data Transport/Storage Question 2)
6. Whether access is centralized or not; in order to provide authorized parties with access to logs, having ICANN be the party that aggregates logs and provides access is the only feasible model. Would require CPs to submit their log data to ICANN on some schedule and in a standardized format. How third parties can access that data is an implementation choice for ICANN. (See Data Transport/Storage Question 3)
7. Benedict/John/Tomofumi to send language on current security/resiliency standards, referencing e.g. ISO standards, etc.
8. All to email list with features they believe are required for the system. To be discussed on 22 Jan. 2019 TSG call.

**Call with CPH (Jan 17, 2019)**

Introduction

1. Goran

2. Ground rules - all calls are recorded and made public, all mailing list discussions are public and archived

3. Selection process

4. Scope of work - charter

5. How we work - cadence, timetable, ICANN Kobe, etc.

6. Fit with EPDP

7. How to provide input into the effort

8. What happens to the output of the group's work

--

**Talking Points**(from Ram):

This system is envisaged to provide access to non-public data to a relatively limited set of authenticated users.

We hope this imposes a gating factor on the scaling of the system or requests.

Access, ie., who asks the question should remain at ICANN

We envision 2 kinds of operational models:

Synchronous

Asynchronous/offline

Explain sync. vs. async. models

==>Goran:

Reason why we are doing this: Current interpretation of GDPR does not make it possible to have a UAM. Contracted parties have legal risk becuase they have the data and collect the data. The law is not made for ICANN or Whois.

If we want a unified way to access, need to find a way to lower risk for CPH. If we don't do it, CPH has to do this on a case and case basis.

<==