Sullivan – a suggestion for "purpose in detail"
http://mm.icann.org/pipermail/gnso-rds-pdp-wg/2017-March/002575.html

I left the meeting with data protection experts last week feeling quite strongly the need for a specific and concrete purpose for each datum we recommend to collect and to make available; and the need for a definition of who the maximal (appropriate) audience is (given the purpose).

At the same time, I think that a reasonably short and high-level statement of purpose along the lines that we have been preparing can provide a useful set of principles.

It strikes me that maybe we could take the high-level purpose statement, and go through some potential data elements and link each one concretely to at least one of the principles in our candidate list.

In what follows I name these "purpose 1", "purpose 2", &c. The purposes are numbered according to the scheme in RDS PDP Phase 1: KeyConceptsDeliberation-WorkingDraft-7March2017 (on p7). I'm aware that the details in the candidate list are still in flux, but I think the broad strokes are pretty close anyway, so I thought I'd try it with the "thin" data we agreed to start with.

This mail is a little long because I'm dealing with all the classes of elements in one message. I suppose we could break this into one-thread-per-element (or class) if we don't converge quickly on each of them. The outline below is just my view, of course, though obviously I think that what I say is true.

I use the "maximal audience" because I think that if there is any "whole public" use then there's no point considering more restrictive uses. (For instance, if we need the domain name to be published to everyone on the Internet because it won't work otherwise, then it makes no difference if LEOs want that data under some sort of authorized-access protocol, because they'll just get it under the wide-open rules instead. So we don't need to care about the LEO purpose in that case.) "Maximal audience" might not work for cases where two different classes have different needs both of which require some restrictions, but it's handy here because we're talking about thin data.

I'm sorry this is long, but I hope it is a useful contribution to the discussion.

Best regards,
A

---%<---cut here---

Here is a convenient example thin whois response, in case anyone wants it to for reference in what follows. (Among other things, it reminds me that something I started to do has never been completed, so thank you to this WG for reminding me of that.)

Domain Name: ANVILWALRUSDEN.COM
Registrar: TUCOWS DOMAINS INC.
Sponsoring Registrar IANA ID: 69
Whois Server: whois.tucows.com
Referral URL: http://www.tucowsdomains.com
Name Server: NS1.SYSTEMDNS.COM
Name Server: NS2.SYSTEMDNS.COM
Name Server: NS3.SYSTEMDNS.COM
Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited
Status: clientUpdateProhibited https://icann.org/epp#clientUpdateProhibited

Sullivan – a suggestion for "purpose in detail"
http://mm.icann.org/pipermail/gnso-rds-pdp-wg/2017-March/002575.html

Updated Date: 17-jan-2017
Creation Date: 30-jun-2010
Expiration Date: 30-jun-2017

1. DOMAIN NAME

a. Collection

The domain name is required to be collected under purpose 1. Without this, there is no domain name, so it is literally impossible to have anything to collect or publish.

b. Publication

The domain name is required to be published under purpose 1, because it is a key by which data is accessed. If you wish to look up the current data about a particular name, you use the name as the key by which you query. (This is not the only possible key. For instance, in an EPP registry you could in principle use the ROID to look up a particular name object. But that does not give you the current data for the thing so named; it just gives you the data about that Repository Object. Two different versions of the same name -- like if example.com is registered by Alice then deleted and later registered by Bob -- have different ROIDs.)

c. Maximal audience

The data audience is Internet-wide under purpose 1 or purpose 2 (or both). The domain name is by definition not private data, because domain names registered in DNS domain name registries (i.e. every registry possibly covered by ICANN policy -- the registries subordinate to the IANA DNS name registries) are name registration in a public name space. Note that it is not possible to keep the existence of a name private, because even if a name were initially undisclosed its existence would be disclosed whenever someone else tried to register it.

2. REGISTRAR IDENTITY

There are four items here, but three classes of data. The (i) registrar ID provides data about the entity that created the entry in the registry (formally, in EPP, "repository"). The (ii) Whois Server and Referral URL both provide metadata necessary for the operation of the distributed database that makes up the RDS (in systems other than whois, approximately the same data with the same relation to identity would be in place, but the details might be different. I think we can treat this as a class anyway). Finally, IANA has a registry of registrar IDs (https://www.iana.org/assignments/registrar-ids/registrar-ids.xhtml#registrar-ids-1), and that contains their (iii) names. This is a protocol parameter registry, but it appears to be managed by ICANN so it is probably appropriate for this PDP to make the policy about how that is to be managed.

a. Collection

Data (i) and (ii) are all required to be collected under purposes 1 and 2. Without this data it is not possible to know the source of the data and it is not possible to trace it further in the system. Data (iii)

needs to be collected in order to give (i) meaning, because it is the only way to know whether two IANA ids are bound to the same organization or person.

b. Publication

Data (i) are possibly required to be published under purpose 1. This largely depends on whether we think the identity of who is managing an object in the registry is part of the "lifecycle of a domain name". My feeling is "yes". Also, this information is likely to be disclosed anyway; see below.

Data (ii) are required to be published under purposes 1 and 2, as long as there is at least one data element that is required under some purpose and is not available from the registry. (Since the actual registration life cycle is controlled by the registrar and not the registry, this appears likely.) Owing to the way these work, publication of these is likely to "leak" information about (i) or (iii) also.

c. Maximal audience

Given purposes 2 and 3 and probably 5, and since the source of contact information is registrars, the maximal audience is probably everyone on the Internet. If we think that purposes 2, 3, or 5 are limited in respect of who needs to make such contact or who needs to check accuracy, then the maximal audience is the set of all those who have such a need. It's worth observing, however, that at least the technical contact for a name ought to be contactable by anyone on the Internet, since when we want to "facilitate communication with domain contacts" at least part of the reason is as a fallback when a site breaks in some way. (This may suggest that we need to unpack the details of purpose 3.)

3. NAME SERVERS

a. Collection

Without collecting the name servers, domain names cannot function on the Internet, so this is required under purposes 1 and 2. (Given that the registration of the name itself and the collection of the name servers are both required for the basic functioning of the Internet Domain Name System, it strikes me that we may be missing a more obvious purpose in our list, but I guess (1) and (2) will be enough and we're already so late that I am loathe to suggest something more.)

b. Publication

Whenever a name is available on the Internet, the name server data is already available in the DNS, so this data is necessarily published. Under either purpose 1 or 2 (or both), the data about nameservers in the RDS provides an avenue for troubleshooting issues in the DNS, and so it is required for those purposes.

c. Maximal audience

Anyone who wants to access a site must be able to find this data in the DNS. Potentially anyone who has a problem with resolution can use the data in the RDS to aid in troubleshooting, so the audience under purpose 1 or 2 (or both) is everyone on the Internet.

Sullivan – a suggestion for "purpose in detail"
http://mm.icann.org/pipermail/gnso-rds-pdp-wg/2017-March/002575.html

4. STATUS VALUES

a. Collection

The status values are not exactly "collected", but are at least in part the result of various actions by the sponsoring registrar and registry on the name. (Some can be set directly.) These govern the disposition of the name in question, and are a necessary condition for having a shared registration system, so they are required under purpose 1.

b. Publication

The status values govern the possible things that could be done to a name, and therefore the data must be published under purpose 1.

c. Maximal audience

At least some status values are required for doing some troubleshooting of resolution failures, so the audience for at least some values under purposes 1 or 2 is "everyone on the Internet". It is possible to argue that some of the status values are relevant only to those people who wish to perform some actions on the domain (such as transferring) or people in a position to do some kinds of activity (such as updating contact information). If we really think it necessary, we could undertake the exercise of audience evaluation for each EPP status.

5. DATES

While the dates might appear to be different kinds, they aren't, since for our purposes they all have at least one common utility (see below).

a. Collection

The dates, like status values, are not exactly "collected": they're a consequence of certain activities. They're necessary for the workings of the shared registration systems using the current fee-for-term model that (approximately?) all gTLD registries use today, so they're required under purpose 1.

b. Publication

The dates are required under purpose 1 or 2 in order to aid troubleshooting of resolution. (If a name worked yesterday and not today, it is helpful to know that it was just created -- meaning the old one was deleted -- or that it is expired, or that someone updated the name only last night.)

c. Maximal audience

Because of the troubleshooting aspects of these dates, the audience under purpose 1 or 2 is everyone on the Internet.

Sullivan – a suggestion for "purpose in detail"
http://mm.icann.org/pipermail/gnso-rds-pdp-wg/2017-March/002575.html

For reference, here are purposes listed in KeyConceptsDeliberation-WorkingDraft-7March2017.pdf

Draft Registration Data and Directory Service Statement of Purpose:

This statement is intended to define the purpose(s) of a potential Registration Directory Service (RDS) for generic top-level domain (gTLD) names. The statement identifies Specific Purposes for registration data and registration directory services.

Note that it is important to make a distinction between the purpose(s) of individual registration data elements[1] versus the purpose(s) of a RDS, i.e., the system that may collect, maintain, and provide or deny access to some or all of those data elements and services related to them, if any.

Specific Purposes for Registration Data and Registration Directory Services

1. *A purpose of gTLD registration data is to provide information about the lifecycle of a domain name.*
2. *A purpose of RDS is to provide an authoritative source of information about, for example, domain contacts[2], domain names and name servers for gTLDs, [based on approved policy].*
3. *A purpose of RDS is to identify domain contacts and facilitate communication with domain contacts associated with generic top-level domain names, [based on approved policy].*
4. *A purpose of gTLD registration data is to provide a record of domain name registrations.*
5. *A purpose of RDS policy is to facilitate the accuracy of gTLD registration data.*

---

[1] Here, "registration data elements" refers to data about generic top-level domain names collected in the relationship between registrars to registries and in the relationship between registrars/registries and ICANN.

[2] Contacts related to the domain name, including those directly related to the domain name and also those involved in the registration system as relevant. Further specification may occur at a later stage in the RDS PDP process.