Merged text from Sullivan's Purposes in Detail & EWG Report Annex D

The EWG Report, Annex D: Purposes and Data Needs, summarizes the RDS data elements recommended in Section IV of the EWG Report, mapped to permissible purposes defined in Section III of the EWG Report. See also Section IV of the EWG Report for collection and disclosure recommendations for each data element. Below appears the portion of Annex D which addresses the "thin" data elements included in the following example WHOIS record:

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

In the table that follows:

Column 1: "Thin" Data Element	Identifies the element by name used in the above example record, followed by the name used in the EWG Report where those differ.
Column 2: EWG Purposes	EWG permissible purposes that use the associated data element. The RDS PDP WG previously reached rough consensus on these as **legitimate purposes for "thin data" collection**; see KeyConceptsDeliberation-WorkingDraft-7March2017.pdf Section 2.2.
Column 3: Collection Rationale	"Collection" text from Sullivan-SuggestionForPurposeInDetail.pdf
Column 4: Publication (Public Disclosure) Rationale	"Publication" text from Sullivan-SuggestionForPurposeInDetail.pdf; appears to refer to what the EWG Report terms "Disclosure" – for "thin" data elements, all public.

For reference, below are the specific purposes for gTLD registration data and directory services (RDS) listed in Section 2.3 of
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.

The wording of each specific purpose is still being refined through WG deliberation and polling.

Columns 3 and 4 in the table that follows refer to purposes 1 and 2 from the above Statement of Purpose.

Thin Data Element	EWG Purposes	Collection Rationale	Publication Rationale
Domain Name	All	The domain name is required to be collected under the Statement of Purpose, purpose 1. Without this, there is no domain name, so it is literally impossible to have anything to collect or publish.	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.)
Registrar	Domain Name Control Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation DNS Transparency	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. Data (iii) needs to be collected in order to give (i) Registrar ID meaning, because it is the only way to know whether two IANA ids are bound to the same organization or person.	See data (i) Registrar ID?

Merged text from Sullivan's Purposes in Detail & EWG Report Annex D

Thin Data Element	EWG Purposes	Collection Rationale	Publication Rationale
Sponsoring Registrar IANA ID (aka Registrar IANA Number)	Domain Name Control Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation DNS Transparency	(i) Registrar ID provides data about the entity that created the entry in the registry (formally, in EPP, "repository"). Data (i) is required to be collected under RDS 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 (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; owing to the way these work, publication of these is likely to "leak" information about (i) and (iii)
Whois Server and Referral URL (aka Registrar URL)	Domain Name Control Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation DNS Transparency	(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) Data (ii) is required to be collected under purposes 1 and 2 (dissemination of registration data). 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 (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.)

Merged text from Sullivan's Purposes in Detail & EWG Report Annex D

Thin Data Element	EWG Purposes	Collection Rationale	Publication Rationale
Name Servers	Domain Name Control Technical Issue Resolution Domain Name Certification Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation	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.)	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.
Statuses (aka Registration Status, Client Status (Registrar) Server Status (Registry))	Domain Name Control Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation	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.	The status values govern the possible things that could be done to a name, and therefore the data must be published under purpose 1.

Thin Data Element	EWG Purposes	Collection Rationale	Publication Rationale
Updated Date and Creation Date and Expiration Date (aka Registrar Expiration Date)	Domain Name Control Business Domain Name Purchase/Sale Academic/Public Interest DNS Research Regulatory/Contractual Enforcement Criminal Investigation/ DNS Abuse Mitigation	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). 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.	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.)

In addition, Sullivan-SuggestionForPurposeInDetail.pdf provides rationale for "Maximal Audience," noting: *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.*

This concept has not yet been included in the above table but can be added in a second pass.

In addition, below are several EWG principles defining criteria for collection and disclosure. These are currently listed in Section 3.4 of
KeyConceptsDeliberation-WorkingDraft-7March2017.pdf

From Pages 41-42:

No.	Data Element Principles
19.	The RDS must accommodate purpose-driven disclosure of data elements. (See Section III [of the EWG Report] for a list of permissible purposes and associated Purpose-Based Contacts (PBCs).)
20.	Not all data collected is to be public; disclosure must depend upon Requestor and Purpose.
21.	Public access to an identified minimum data set must be made available, including PBC data published expressly to facilitate communication for this purpose.
22.	Data Elements determined to be more sensitive (after conducting the risk & impact assessment) must be protected by gated access, based upon: • Identification of a permissible purpose • Disclosure of requestor/purpose • Auditing/Compliance to ensure that gated access is not abused
23.	Only the data elements permissible for the declared purpose must be disclosed (i.e., returned in responses or searched by Reverse and WhoWas queries).
24.	The only data elements that must be collected are those with at least one permissible purpose.
25.	Each data element must be associated with a set of permissible purposes. • An initial set of acceptable uses, permissible purposes, and data element needs are identified by [the EWG] report (see Section III and Annex D). • Each permissible purpose must be associated with clearly-defined data element access and use policies. • As specified in Section III, an on-going review process must be defined to consider proposed new purposes and periodically update permissible purposes to reflect approved additions, mapping them to existing data elements. • A Policy Definition process must be defined to consider proposed new data elements and, when necessary, update defined data elements, mapping them to existing permissible purposes.
26.	The list of minimum data elements to be collected, stored and disclosed must be based on known use cases (reflected in [the list of permissible purposes]) and a risk assessment (to be completed prior to RDS implementation).

See also Data Collection and Data Disclosure Principles (Pages 42-46)