

Comments on the Categories for the Triage Document

I commented during the call on July 20/21 that the categories chosen by Susan and Lisa to group our requirements, appear to be somewhat random. Chuck asked me to suggest a different way of doing this. This note is an attempt to both better explain my objections, and provide a more analytical framework for sorting and winnowing our possible requirements.

First, let me reiterate that this is a tremendous amount of work, and thank Susan and Lisa for doing it. As indicated during the call last night, this is a simple, hopefully neutrally phrased recognition, that we all bring our particular experience, vocabulary and expertise to this exercise. We all view things differently, and thus if one person chooses the categories/keywords it will in all likelihood not necessarily reflect the words and categories that other stakeholders hold dear.

Stephanie Perrin

Why is this important?

We have just gone through an extensive exercise of locating documents we thought relevant to the exercise. This was done uncritically, all docs accepted. If there were more documents in a particular category or interest of some stakeholders, we did not weight the category differently. Similarly, we did not categorize the documents with respect to their authority....there were wish lists, letters, opinions of judicially authorized officials, early documents and discussion papers from the initiation of ICANN, opinions of ICANN's SSAC, the actual legal requirements of registrars (RAA) etc. There is a great disparity in the status of these documents which we have yet to sort out in our deliberations.

Next, we sought volunteers to summarize these documents. A herculean effort, kudos to all who did this, I am still late in a couple of my tasks. I looked at some but not all of these summaries, comparing them to the originals, for documents that I am extremely familiar with, and for some that I am less familiar with and are not in my area of expertise. I feel safe in saying a great job was done by all, but they vary in emphasis, style, vocabulary (i.e. were they summarized in the language of the person doing the summary or did that person repeat the language of the document?) and thoroughness. Confident that Chuck means what he says when he says we can revisit this when we deliberate, I did not start quibbling about the summaries. We need to get on with our work, as everyone has said.

However, now that we are going back and pulling out keywords to sort the requirements, the actual language of the requirements becomes much more important. In my view, while recognizing that there has been a lot of work done already on the triage document, it will be easier to install a neutral frame on the

sorting mechanism (the keywords, groups, categories, however you choose to describe them) than to go back and start kibitzing about the summaries. I find it hard to understand how the categories in which we group requirements are not important. I am making an honest effort to explain my concerns, if folks who feel that how we categorize is not important can explain to me how we are not going to bias our work, I would be delighted to relax about this. Really. I have circulated this draft to my colleagues in NCSG, where we have several academics who are expert in qualitative analysis, and they concur that even if this is simply a rough and ready sorting tool, the framework is important.

Here are a few of my concerns:

1. Most people on this group are not going to read all the documents. I am reading most of these documents for my doctoral dissertation, and yet I had not read everything we came up with. I cannot remember the content of many of them, I have to go back to my notes. I think it is reasonable to conclude we will be referring only to the summaries and requirements, and therefore the language used in them, the emphasis becomes important. Obviously, folks like me will be dragging people back to the originals, but I am familiar with how grumpy people get in lengthy deliberations; there will be little patience for this as time goes by. Gaps may linger indefinitely.
2. Some areas of expertise are better understood at ICANN than others. Privacy and data protection, human rights and rule of law, I would suggest, are areas where many people have less expertise and possibly less interest. This is my area, naturally I care about how certain things are dealt with. I don't lose sleep over whether the registrars will be able to straighten our group out about how such things as the use case Michele went through July 20 actually occur. Lots of expertise there, and the players are engaged. It will be harder to discuss the actual brass tacks of data protection and constitutional protections across borders. Reading the Schrems decision (on our document list) does not mean you instantly understand it, any more than I am ready to join the IETF after I plow through the RDAP technical documents. I fear the groupings we propose will be taken by those unfamiliar with the various fields as representative of the key principles or elements. At the moment, they are not. This introduces bias.
3. Let me clear up a fundamental area of confusion. Michele intervened to compare these groupings to the Dewey Decimal system. We can go to a library and quibble about whether a book is in history or in sociology, but you can still find the book. True. Especially if you have a librarian to find it for you, and you know what you are looking for. I would suggest (see 2 above) that with respect to many topics we cover, not all of us know what we are looking for. That spreadsheet becomes our finding aid. Nevertheless, I am not quibbling about how the possible requirements are categorized (yet) I am concerned about the categories in the Dewey Decimal system. Is history a good category? Do we agree that biography is a good category? Sociology? I think while the system is not perfect, much thought went into how to frame the key groups. Let me give an example: Cookbooks.

If I am browsing cookbooks, I will see dessert, bread, regional (eg. Mediterranean), etc. all subcategories of cookbooks. Included are health related ones. Now the health ones could have been categorized as related to each particular condition (weightloss, diabetic, gluten free, etc) in the Medicine section, and a good argument could be made that everything related to health belongs in whatever other medical category exists, or age categories (eg. making baby food, feeding athletes etc.). The decision was made that a cookbook is a cookbook, contains recipes, etc. Why is it better to have all cooking related material together? One of the purposes of a library is to help people learn. Attention is paid to ensuring homologous items are together. Perhaps while checking out the weight loss cookbook I will pick up the diabetic one and learn more about impacts of caloric choices....

None of these examples are perfectI am trying to illustrate the overall point that agreeing a logical sorting framework is important for many reasons. Parking certain possible requirements in odd categories will bias the search results of our members. Some things will not show up as frequently as they should. Some may not be found, making cross-referencing incomplete.

4. We are now proceeding with the task of identifying dependencies, and the particular phase of the workplan when these possible requirements will be discussed. I suspect that we will be living with this spreadsheet for a long time as we work through the workplan. Chuck has assured us repeatedly that we can revisit all aspects of this work in the deliberation phases. He has not assured us he will be there four years from now. This makes me nervous, as I envisage the prospects of pointing out the inadequacies of a summary, the lack of completeness of a requirement list, or the mis-categorization of a given item. Furthermore, it is not a given that any of us are going to be here till the end of the process. I think it therefore important to make sure we get this sorting mechanism right. Most of us are so delighted that Susan and Lisa did all this work we are not likely to critique it, but we need to focus on and agree the framework for sorting possible requirements.

Current Frameworks to sort our Possible Requirements

As discussed above, the main points of each document were pulled out in a rather random way. We have already sorted the documents as to whether they address the questions of our Charter:

- FQ Foundational Questions: Questions to be answered based on all other requirements
- OQ Other Questions: Questions that may not fit within the 11 charter questions
- UP Users/Purposes: Who should have access to gTLD registration data and why?
- GA Gated Access: What steps should be taken to control data access for each user/purpose?
- DA Data Accuracy: What steps should be taken to improve data accuracy?

- DE Data Elements: What data should be collected, stored, and disclosed?
- PR Privacy: What steps are needed to protect data and privacy?
- CX Coexistence: What steps should be taken to enable coexistence?
- CM Compliance: What steps are needed to enforce these policies?
- SM System Model: What system requirements must be satisfied by any implementation?
- CS Cost: What costs will be incurred and how must they be covered?
- BE Benefits: What benefits will be achieved and how will they be measured?
- RI Risks: What risks do stakeholders face and how will they be reconciled?

However there is a great deal of overlap, and we did not quibble too much about whether each document was tagged appropriately. We might have wound up with many documents in all categories. So far, we have focused on three categories, DE, UP, and PR.

In many ways, sorting for the purposes of the spreadsheet is another transverse slice at sorting. If we examine the current buckets, we can detect a certain logic to them, but it is not consistent. I have inserted my questions below.

CODE	DESCRIPTION	COMMENTS
A	Purpose	
B	Contact data for technical resolution	this is a fine grained element under data elements, important but not a category in itself
C	Registration data query, search and disclosure	This is a function that RDS users want
D	Policy needs	Unclear what this bucket means. Is it a big bucket for all policy considerations?
E	Identifying own data and access	Does this refer to end user data access rights under DP law? Or input validation?
F	Contact data for other than technical resolution	Again, a fine-grained data element issue
G	Proxy	Does this refer to a data element that identifies whether an end user has employed a proxy service, or is it a big bucket for all things proxy?
H	Extensibility	Is this one element in large bucket of technical requirements, or is it a

CODE	DESCRIPTION	COMMENTS
		particular function we need to draw out separately (and if so why)
I	Research (other than for legal investigation)	This is one potential use of RDS data. If we pull it out, do we need to pull out all potential uses?
J	Legal research	If this refers to the use of RDS data for legal research, see I above. At present it could also refer to legal research on applicable law, as we have all kinds of documents discussing legal interpretation, mostly of Data Protection law.
K	Registrar transfer policy	Why is this policy pulled out separately? Why not the WHOIS conflicts with law policy, or any other policy? Surely all policies must be sub categories of the larger policy bucket?
L	Consent	This is one aspect of data protection law in some jurisdictions, and it is a highly problematic one. I would disagree with emphasizing it in isolation of other relevant principles/requirements.
M	Controller/Processor/Processing or transfer of data	These terms are all present in European data protection law. They are defined terms and relevant in the law, but cannot really be pulled out in isolation. Furthermore, it is an aspect of the contract between the Registrar and the Registrant, which begs a bucket on contracts, and this one is not under ICANN control except insofar as the RAA provides requirements.
N	Accuracy of data	Accuracy of data is a good example of a bucket that applies in various activity areas, including compliance, law enforcement investigations, data protection law, etc. Attention to accuracy of data could be considered a function bucket, accuracy itself a policy goal.
O	Retention of data	Retention of data is a function required by the RAA. Should some of these functions be grouped?
P	Use of data for surveillance	Surveillance of what? Individuals and companies do surveillance of data for marketing, cybercrime enforcement,

CODE	DESCRIPTION	COMMENTS
		compliance, competitive intelligence, criminal investigation, risk management, etc. Is it a use or a function?
Q	Law Enforcement Investigation	As has been pointed out on the list, the vast majority of investigation using RDS data is done by private actors. I would suggest a general term investigation, then sub categories.
R	Proportionality of use of the data	Proportionality is an important principle in EU law. Data protection is evaluated on the basis of that principle. There are others that are also important, this is a subset of basic human rights law/obligations and is phrased differently in different jurisdictions. Yes it appears in many documents, because of our reliance on EU texts for data protection and human rights analysis, but it is indeed a subset of a broader category of rights.
S	Gated Data Access	This is a recommendation of the EWG report. It is also one of our sorts for the documents, already, so is somewhat duplicative as a bucket here.
T	Public Data Access	This is a subset of gated data access (ie the top category), if looking at the term through the lens of the EWG. It is the current condition if looking through the lens of WHOIS. It is a function that is a potential requirement of the new RDS, so I would be tempted to put it as a subcategory of a functions bucket rather than pull it out separately.
U	Access Policies, including Authenticated Access	Authenticated access is a technical requirement for gated access (among other internal roles). Access policies are required to set the frame for those technical requirements, as has been discussed recently on the list. I would suggest this belongs in the policy bucket.
V	Access Violation	There is no bucket here called abuse. Access violation is a kind of abuse, but there are many others. It flows from the concept of gated access, but logically belongs in its own bucket as a type of abuse, in my view. One could also view it as a security issue and establish a security

CODE	DESCRIPTION	COMMENTS
		controls bucket, but I think perhaps abuse might be more useful (certainly a debatable point).
X	Encryption	Encryption is a technique to ensure security. There are many others. We do not have a security bucket. Perhaps we need one (see v above).
Y	Internalization	Beats me what this means or comes from. I have not been through the spreadsheet yet in detail, but I don't think it is obvious from the keyword, too fine grained?
Z	Audit or Logging	Audit or logging is a function of the system, and/or a function of policy requirements. It could be classified as a security technique, or a compliance technique.
NOTE	I would recommend that the following categories be changed to relate to the bucket of which they are a subset.....at present, A is purpose, B is contact data for technical matters, and it is quite confusing to reuse the initial letter again in this way to apply to unrelated items	
Aa	Validation of Contact Data	Subset of accuracy or compliance
Ab	Applicable Law	This is a broad category that should go above. Covers all kinds of law, plus jurisdictional matters, and many of our documents have requirements that relate to this complex issue.
Ac	Cert Authority (or any third party that has duty to validate)	I would agree this function (Cert management) is important but I don't see that it needs to be a category, wont help us sort our requirements. If it is anything, it belongs as a function, not an individual or actor as it appears here.
Ad	Transparency	A broad category that relates to data protection law, goals of the system that are in keeping with ICANN's public good mandate etc. Suggest it as a sub category of policy, as a general policy goal.
Ae	Validators	Similar to Cert authority. Validation is a function that we may wish to bucket, validators is a fine grained determination as to how to do that (belongs in implementation and we cannot pull out

CODE	DESCRIPTION	COMMENTS
		every actor for every function).
Af	Contact validation and agreement	Contact validation is a function. Not sure which sense of "agreement" we are talking about; if it is an instrument, need to describe what kind of agreement/contract we are talking about, but if it is in the sense of a function, is it the technical agreement/coherence of contact information across jurisdictions, systems, and languages? Need clarification.
Ag	Cost	We have cost as a sort above in CS. Agree that as a practical consideration cost is important, but as Chuck pointed out a while ago it is an implementation/design issue. Perhaps we need a bucket of implementation issues in which cost is a subset?
Ah	Unique Contact data	See F and B above. This is one of the things that has me concerned about this list, we have several buckets, all relating to contact data. Contactability is a function or feature of the system, needs to be one bucket with subsets
Ai	Synchronized	Unclear what this means, assume technical requirement.
Aj	Authoritative data	Is this a system requirement, a legal requirement, a law enforcement requirement, a choice in terms of data distribution and reliability?not clear. Lets figure out the function and drop documents discussing the issue of authoritative data into the relevant bucket.
Ak	Stability data elements	Too fine grained, is it a technical requirement? If so, lets put it in the technical requirement bucket, if it is a cluster of data elements representing (and collected to provide) the attribute of stability, we need that explanation.
Al	Display	Are we talking about data display in a potential RDS? Why pull this out separately, does it relate to technical requirements, policy, or data protection requirements?
Am	format	Technical requirement I presume?

CODE	DESCRIPTION	COMMENTS
An	Account information not data elements	What does this mean? Are we talking about customer information retained by registrars according to requirements of the RAA, but not available in the RDS?
Ao	Registrar Data Elements	Too fine grained, suggest as a subset of a larger bucket although we already have a category of DE.
Ap	Nameserver Data	See Ao.
Aq	RPM	I presume this refers to rights protection mechanisms. Perhaps this deserves a bucket of its own? Rather an important function of what we are doing (i.e. one use of RDS data is to facilitate the smooth functioning of the RPMs which are outside the scope of this debate yet depend upon it. Similar to Law Enforcement Access in that respect.
Ar	Registrar Access	I presume this refers to a function required of the system, namely that registrars be able to access the RDS. Not sure why this has its own bucket.
As	Privacy/control of data	We have a sort above (PR). Any bucket we decide has to serve a function. Control of data is much broader than privacy, it relates to the maintenance of confidentiality and security. Suggest renaming one bucket as confidentiality, that would cover a number of confidentiality issues which could be pulled out as subsets as required.
At	Notice to data owner	Fine grained, again. What notices are we talking about here, and how do we define "owner"? In data protection terms, we avoid this term. Registrars may correctly claim to be data owners for customer data. Suggest dropping it. If this is a contractual requirement under the RAA or under RPMs, then it belongs under contract bucket or RPM bucket. If a compliance term, under compliance.
Au	Accountability – use of data.	Accountability is a broad policy goal, suggest putting it there. If you wish to describe the function of accountability mechanisms, then that perhaps is a bucket under functions. Ambiguous as it stands.

CODE	DESCRIPTION	COMMENTS
Av	Terms and Conditions	Of what? Contracts bucket, I think.
Ax	Data geographical locations	This is fine grained. Is it a technical requirement? Is it a desired element of potential end users (eg. LEAs)?
Ay	Contract	This could be a broad bucket called contractual issues. There are many contracts we could be talking about here....if the RAA we should say so. If contracts for bulk access we need to pull them out separately.
Az	Responsibility for use of the domain name	Again, is this a broad policy goal of the DNS? If so, lets put it under policy bucket.
Ba	Right to privacy	The right to privacy is covered under PR, and refers to rights under data protection law, under common law, under specific statute and under constitutional protections. Suggest a broad single letter bucket, or subset of law.
Bb	Aggregated Data	Not sure what this refers to.

Possible New Framework of Categories

I think it would be better to sort by broad descriptors that refer to whether a series of possible requirements are functions, goals, legal requirements, technical requirements, etc. Taking the existing data and replacing them with broad categories, and using the two letter categories for subsets, I came up with the following new table. I have inserted the existing categories where it seemed to me they logically belonged, and removed categories that seemed too fine grained (these are all debatable distinctions). I have suggested a new field for Keywords, in which people can insert keywords they want to sort by, because we all speak of different themes in our own terminology. This will permit searching by some of the terms already pulled out as categories, but which seem too fine grained. Other keywords can be added. I am not an excel person, so I leave it to Lisa to say whether a new category in the triage document would be useful here, or whether we turn the Key into a finding aid by adding searchable keywords. I would favour the latter because big spreadsheets overwhelm me, but I defer to the group.

CODE	NAME	DESCRIPTION	EXISTING GROUPS TO INCLUDE	KEYWORDS	COMMENTS
A	Goals of System				
AA	Transparency		Ad		
AB	Differentiated access		S, P, U, Ar		
AC	Authoritative Data		Aj		
AD	Accountability		Au, Az		
B	Functions				
BA	Search & Query		C		
BB	Certification, Authorization		Ac		
BC	Compliance				
C	Potential use of data		Bb		
CA	Research		I, J		
CB	Surveillance		P		
CC	Investigation		Q		
D	Data elements		Ao, Ap, Ax		
DA	Contactability		B, F, Af, Ah		
DB	Accuracy		N, Aa, Ae		
E	Legal Requirements				
EA	Privacy, DP law		E, L, M, R, Ba		
EB	IP law				
EC	Jurisdiction Issues		Aa		
ED	Free Expression				
F	Technical Requirements		H, Ai, Ak, Al, Am		
G	Security Requirements		X,		
H	Confidentiality requirements, issues		As		
I	ICANN Policies				
IA	Policy Issues		D, At		
IB	Registrar Transfer		K		
IC	RPM		Aq		
ID	Privacy/Proxy		G		
J	Contractual Issues		O, An, Av, Ay		
K	Implementation Issues				

CODE	NAME	DESCRIPTION	EXISTING GROUPS TO INCLUDE	KEYWORDS	COMMENTS
KA	Cost		Ag		
L	Abuse & Mitigations		V, Z		
M	Risk				
MA	Registrants'				
MB	Trademark Owners'				
MC	Contracted Parties				