

# Preliminary Review of Service Level Expectation Data

July 2016

In developing the IANA Stewardship Transition Proposal on naming-related functions, a design team was convened and developed a set of “Service Level Expectations”. These includes categories of measurement that should be instrumented for the root zone management component of the IANA naming functions, and against which formal service level thresholds should be considered to be assessed.

A number of principles were defined as part of this work. This included that “the definition of specific thresholds for performance criteria should be set based on analysis of actual data. This may require first the definition of a metric, a period of data collection, and later analysis by the community before defining the threshold.”

Further consultation with the community groups lead to ICANN instrumenting its Root Zone Management System (RZMS) to commence capturing the bulk of the measurements in March 2016. This review of approximately three months of data provides a preliminary basis for a first round of threshold definitions.

It should be recognized that while this review is based on early data, the overall post-transition environment envisages regular review and revision by the newly-formed Customer Standing Committee (CSC) of these thresholds. Future work in this area will have more data upon which to rely on to make analyses, along with a better understanding of the post-transition environment.

## Approach to draft thresholds

For each of the performance time measurements, a preliminary data set was obtained that spanned from 2 March 2016 (when a revised version of RZMS was deployed that instrumented events associated with the start and end times of these various measures), and 23 June 2016.

Each data set was individually reviewed and plotted, and a number of basic statistical measures were derived. Statistical outliers were manually reviewed by staff to confirm the accuracy of the measurement.

The overarching direction from the community on setting the performance targets was to maintain “status quo” throughout the transition, that current performance of the root zone management functions is considered satisfactory and should be maintained. Given the nature of root zone management is a very low number of requests, and a high degree of variability of unique circumstances for many of them, our starting approach therefore was to define thresholds for which we have an appropriate confidence level.

Due to the short period of data collection, there are a low number of overall requests from which to derive performance measurements. The majority of the measurement categories for which SLEs are proposed had either zero requests during the collection period, or a small number from which conclusions cannot be satisfactorily derived.

No performance measurements exceeded a count of 100 for a given month which meant thresholds intended to be met at high levels (i.e. 99%) seemed inappropriate — otherwise a single outlier in a given month would mean PTI would fail to meet the service level. Also, given measurements are over calendar days, not work days, certain measurements can be heavily skewed by the day of submission unless the population size is large to average out this effect. Therefore no performance time SLAs have been proposed to a percentage higher than 95%.

For availability measurements, availability for all systems under measurement was 100% during the entire period except for scheduled routine maintenance of the RZM system.

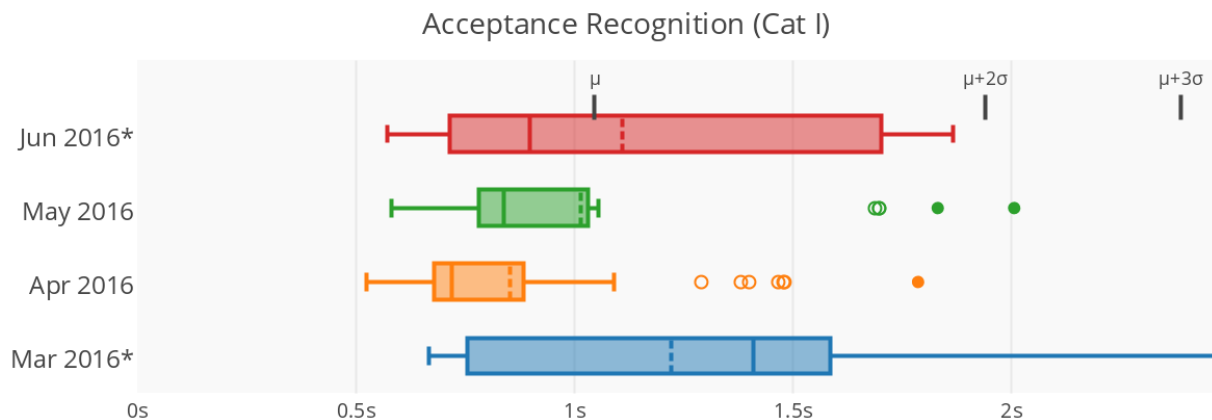
For accuracy measurements, we believe acceptable performance is nothing less than 100% accuracy. All of these measures are therefore proposed to this level without exception.

## Processing Performance (Submission)

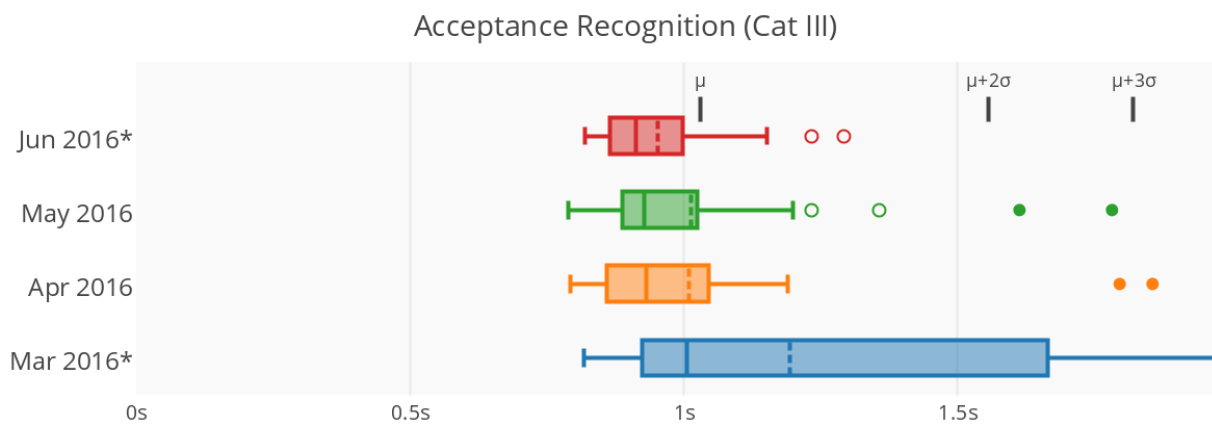
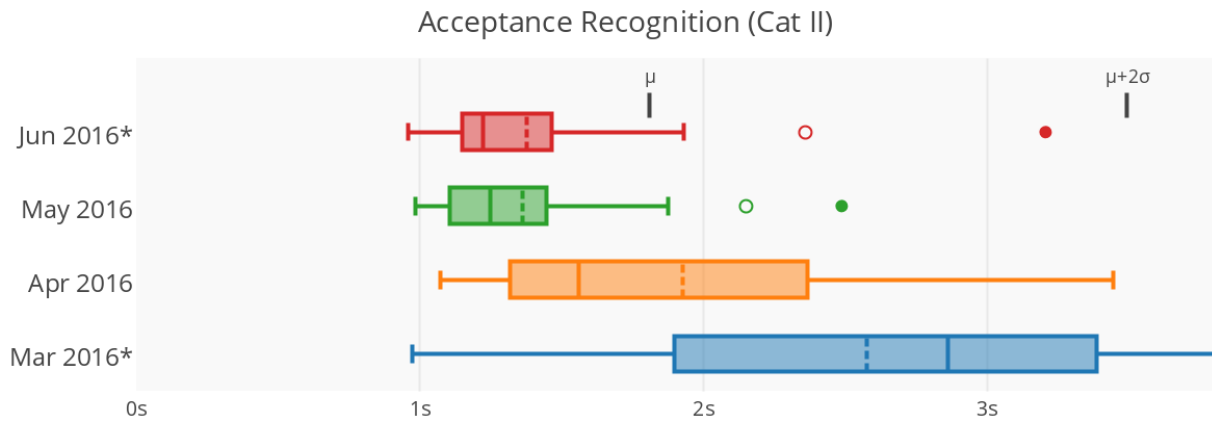
### Time for ticket confirmation to be sent to requester following receipt of change request via automated submission interface

Measurements suggest this performance process typically takes a few seconds to process. Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Category <sup>1</sup>	Data	Recommendation
I	n=142, mean=1.05s, min=0.53s, max=2.48s, pstdev=0.45s, +2s=1.94s, +3s=2.39s	≤ 60s (95%)
II	n=96, mean=1.81s, min=0.96s, max=3.81s, pstdev=0.84s, +2s=3.49s, +3s=4.33s	"
III	n=111, mean=1.03s, min=0.79s, max=1.98s, pstdev=0.26s, +2s=1.56s, +3s=1.82s	"
IV	n=0	"
V	n=1, mean/min/max=0.93s	"



<sup>1</sup> Categories are those as defined in the design team's recommendations. In brief, Category I refers to routine changes that impact the root zone; Category II are routine changes that do not impact the root zone; Category III are changes pertaining to creating or transferring a gTLD; Category IV are changes pertaining to creating or transferring a ccTLD; Category V is all other changes, including those in Cats I-IV that had exceptional processing issues that made it impossible for them to be processed routinely.



## Time for lodgment of change request into RZMS by ICANN staff on behalf of request sent by email

There is insufficient data to assess how long this process typically takes (the process was never invoked during the preliminary measurement period). Setting a service level expectation should be deferred until there is data to analyze and/or a dialogue between PTI and the CSC on determining customer expectation.

Category	Data	Recommendation
I	n=0	—
II	n=0	—
III	n=0	—
IV	n=0	—
V	n=0	—

## Processing Performance (Technical Checks)

### Time to return results for technical checks following submission of request via automated submission interface

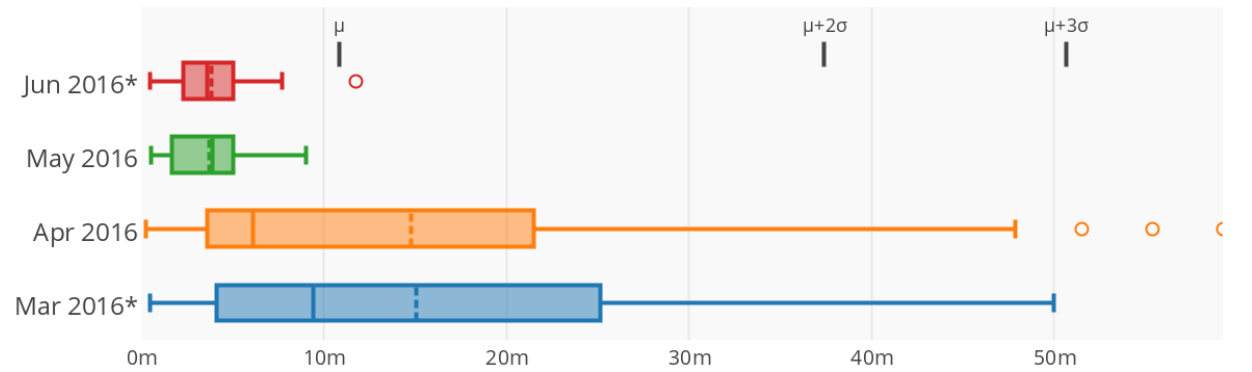
While performance of a technical test is considered IANA processing time, the time taken to execute significantly reflects the configuration of the customer's services. For example, if a customer supplies servers that have connectivity problems, we must probe them, wait for their response and perform a number of retries. These delays waiting for customer responses appear as IANA processing time in the context of a single performance of a technical test suite, thus the time taken is reflective of the quality of the technical configuration provided by individual customers. Given single entities like registry service providers often submit changes in bulk to a portfolio of top-level domains, an issue with their configuration can have a compounding effect for a given period.

The superior performance of Category III tests is expected, as it reflects improved readiness of new gTLD delegations to having their technical configurations conform with the root zone requirements that are tested. This is the result of prior to submitting a Category III request, as part of their new gTLD application with ICANN, the customer must complete *pre-delegation testing* (PDT). This process is comprised of a superset of the IANA root zone technical tests, and therefore, the customer will have remedied any issues we are likely to identify just prior to submission of the request, resulting in the vast majority passing the tests without issue. This dynamic may change in the future as gTLD transfers become the predominant request type in this category, and those types of changes are not preceded by the same PDT regimen.

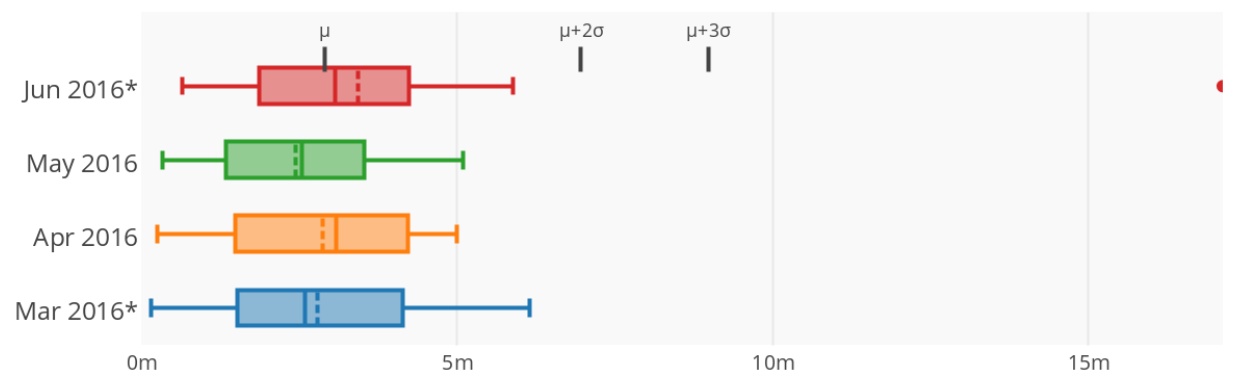
It should also be noted that this is measured as defined by the SLE document, which means that unlike the other technical check measures, this process is not strictly limited to measuring how long the technical check performance process takes. It measures from the *beginning of the request* until the first technical check result is returned, which includes other automated processing including the time measured earlier in the category "Time for ticket confirmation to be sent to requester following receipt of change request via automated submission interface".

Category	Data	Recommendation
I	n=142, mean=10.86m, min=0.28m, max=59.25m, pstdev=13.28m, +2s=37.42m, +3s=50.70m	≤ 50m (95%)
II	Not applicable.	—
III	n=111, mean=2.91m, min=0.16m, max=17.13m, pstdev=2.02m, +2s=6.96m, +3s=8.98m	≤ 10m (95%)
IV	n=0	≤ 50m (95%)
V	n=1, mean/min/max=133.80s	"

Technical Test (First) (Cat I)



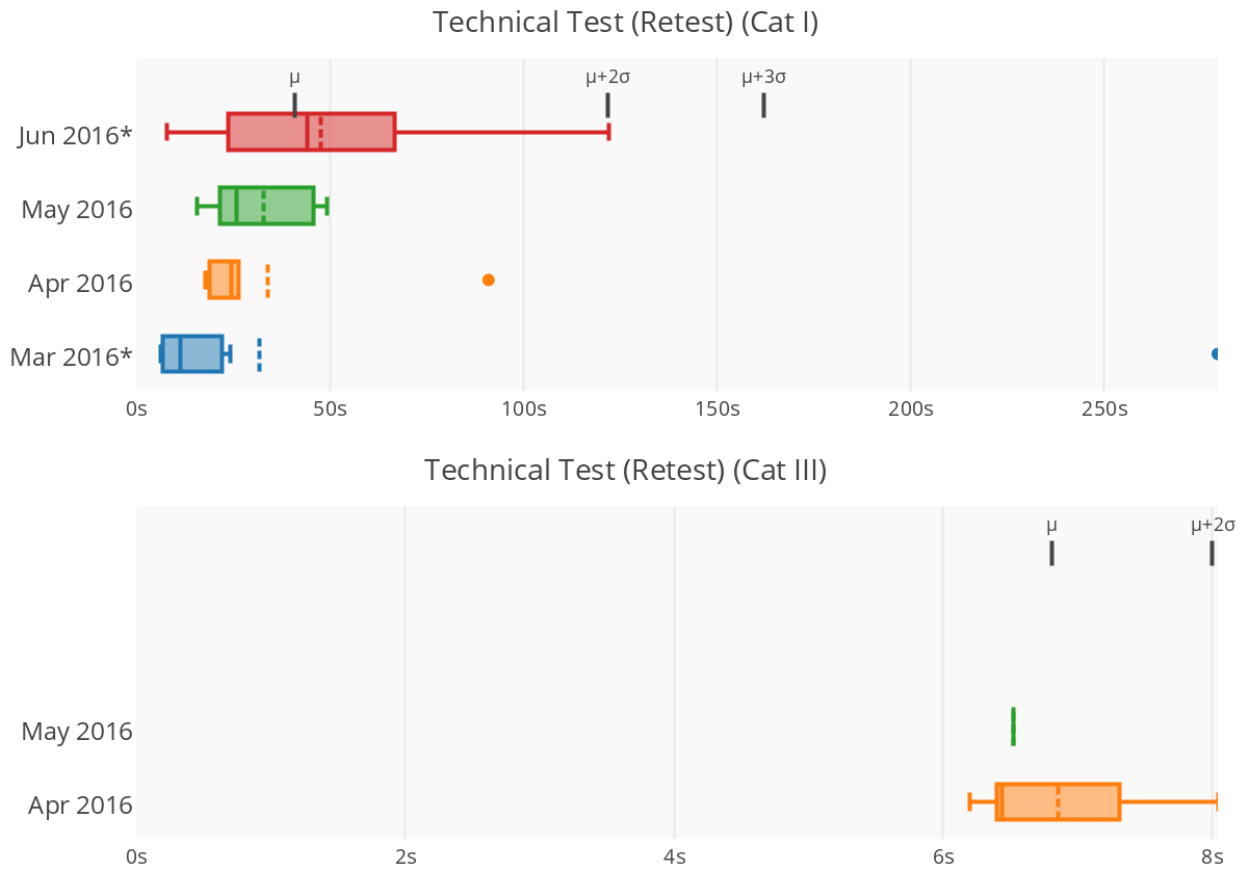
Technical Test (First) (Cat III)



**Time to return results for subsequent performance of technical checks during retesting due to earlier failed tests**

Should a customer not pass technical testing, the system will automatically retry testing periodically to identify if the identified faults have been remedied. In addition, the customer may explicitly force a retest through their self-service interface. This measurement identifies the length of time each of those technical check retest runs takes.

Category	Data	Recommendation
I	n=64, mean=40.84s, min=6.29s, max=279.34s, pstdev=40.44s, +2s=121.71s, +3s=162.15s	≤ 3m (95%)
II	Not applicable.	—
III	n=8, mean=6.81s, min=6.20s, max=8.04s, pstdev=0.60s, +2s=8.01s, +3s=8.60s	≤ 3m (95%)
IV	n=0	"
V	n=1, mean/min/max=4.70s	"



### Time to return results for performance of technical checks during Supplemental Technical Check phase

After the bulk of processing, just prior to releasing a request for implementation, the suite of technical tests is reperformed. This is known as the “supplemental technical check”, the purpose of which is to identify any configurations that fell out of compliance during the IANA processing time. While most requests only take a couple of days to process, this is not normally a concern, but for requests in Categories IV and V, they may take an extended period of many months, and thus the technical environment may have significantly deviated from the one tested at the commence of the ticket. This supplemental check is designed to identify these cases.

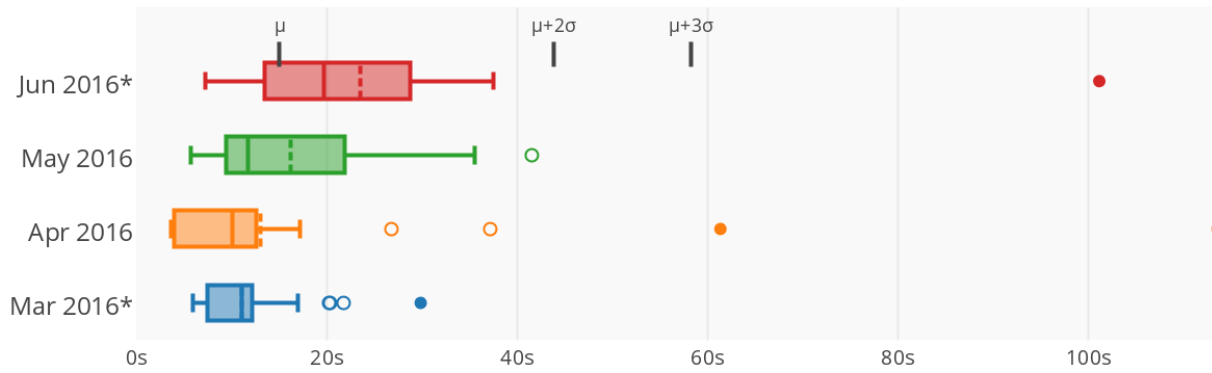
The higher performance of this phase compared to the first technical check and technical check retests is to be expected as a requests typically<sup>2</sup> do not reach this phase without any technical check issues being remedied earlier in processing.

<sup>2</sup> In the event staff processing an individual request grant an exception/waiver to a particular technical requirement due to unique circumstances, the request will still fail the automated supplemental technical check phase. This could cause variance on individual requests which are not exhibited in this preliminary data set.

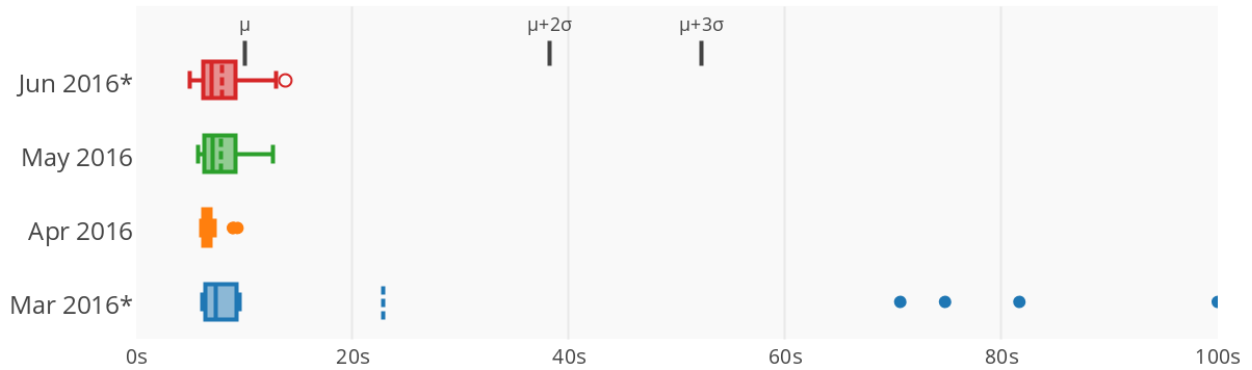
We recommend a higher threshold of 5 minutes for this measure due to the impact of customer configurations.

Category	Data	Recommendation
I	n=142, mean=15.04s, min=3.64s, max=113.65s, pstdev=14.41s, +2s=43.86s, +3s=58.28s	≤ 60s (95%)
II	Not applicable.	—
III	n=112, mean=10.09s, min=4.95s, max=100.04s, pstdev=14.07s, +2s=38.22s, +3s=52.29s	≤ 60s (95%)
IV	n=0	≤ 5m (95%)
V	n=1, mean/min/max=4.40s	"

Technical Tech Check Supplemental (Cat I)



Technical Tech Check Supplemental (Cat III)



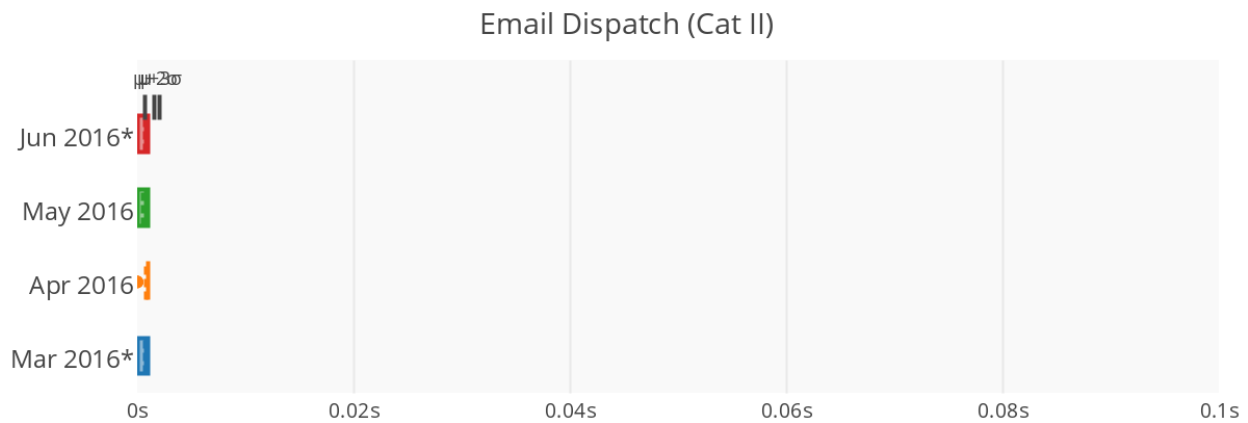
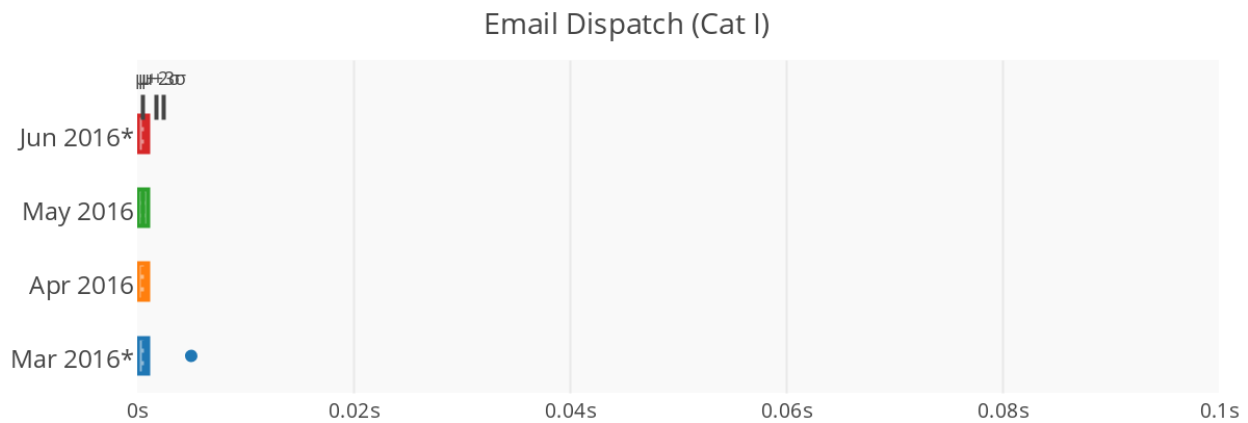


## Processing Performance (Contact Confirmation)

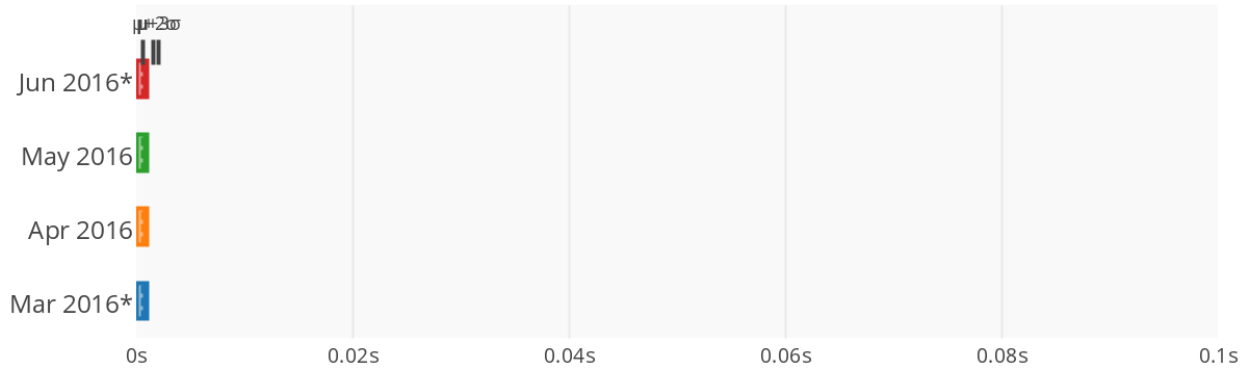
### Time for authorization contacts to be asked to approve change request after completing previous process phase

Measurements suggest this performance process typically takes a few milliseconds to process. Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Category	Data	Recommendation
I	n=134, mean=0.00s, min=0.00s, max=0.01s pstdev=0.00s, +2s=0.00s, +3s=0.00s	≤ 60s (95%)
II	n=96, mean=0.00s, min=0.00s, max=0.00s pstdev=0.00s, +2s=0.00s, +3s=0.00s	"
III	n=107, mean=0.00s, min=0.00s, max=0.00s pstdev=0.00s, +2s=0.00s, +3s=0.00s	"
IV	n=0	"
V	n=1, mean/min/max=0.00s	"



### Email Dispatch (Cat III)

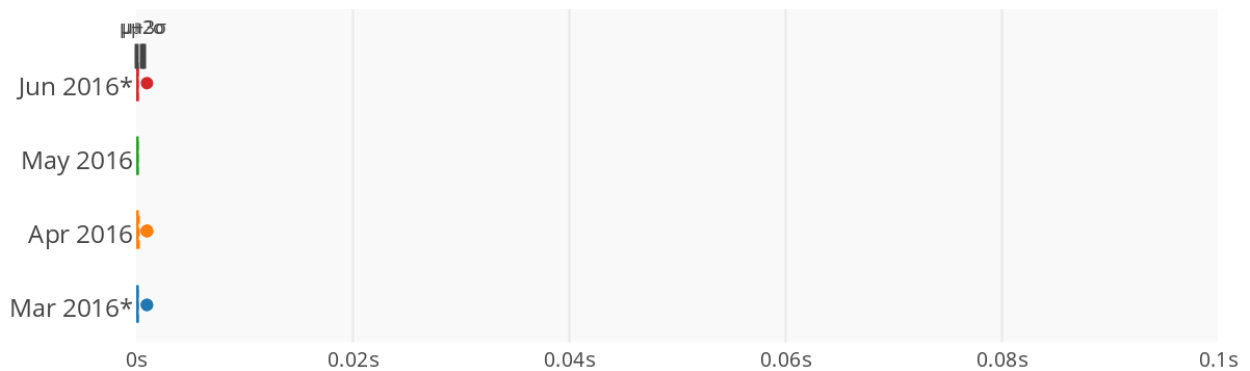


### Time for response to be affirmed by IANA Functions Operator

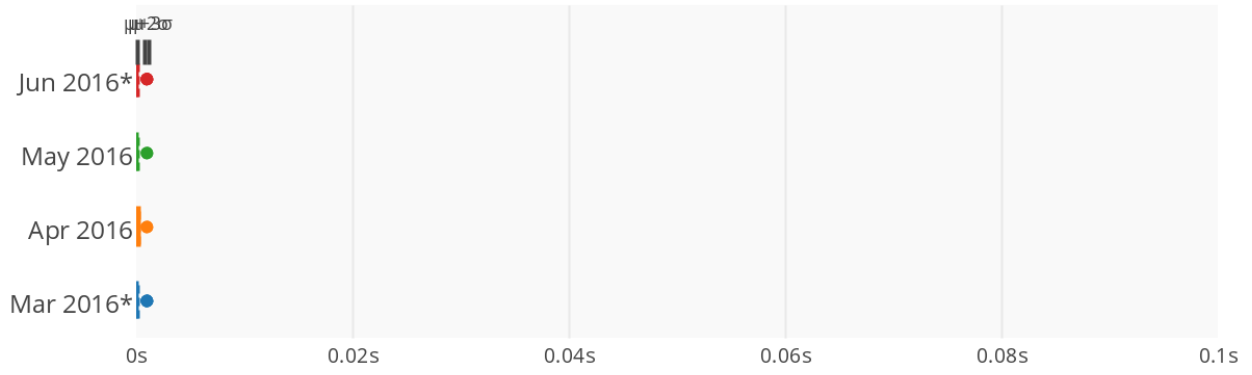
Measurements suggest this performance process typically takes a few milliseconds to process. Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Category	Data	Recommendation
I	n=130, mean=0.00s, min=0.00s, max=0.00s pstdev=0.00s, +2s=0.00s, +3s=0.00s	≤ 60s (95%)
II	n=86, mean=0.00s, min=0.00s, max=0.00s pstdev=0.00s, +2s=0.00s, +3s=0.00s	"
III	n=103, mean=0.00s, min=0.00s, max=0.00s pstdev=0.00s, +2s=0.00s, +3s=0.00s	"
IV	n=0	"
V	n=1, mean/min/max=0.00s	"

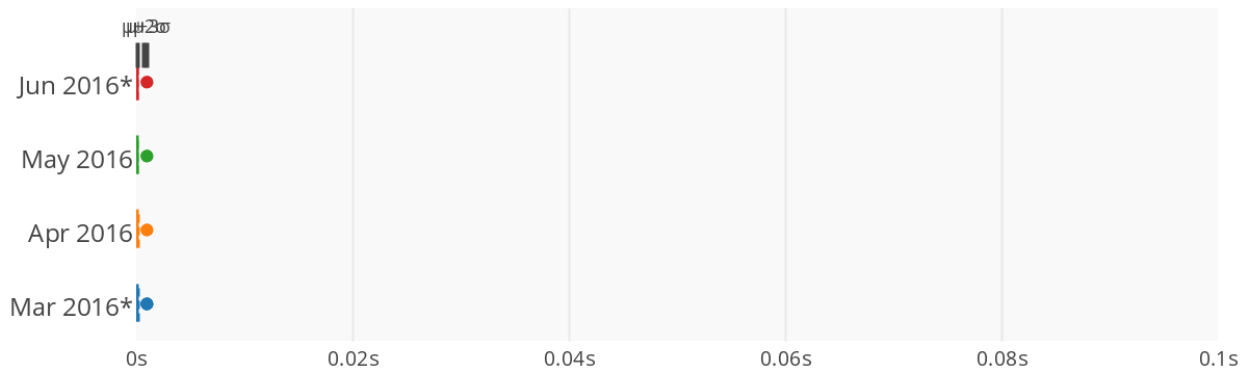
### Recognition of Confirmation (Cat I)



Recognition of Confirmation (Cat II)



Recognition of Confirmation (Cat III)



## Processing Performance (IANA Processing)

### Time to complete all other validations and reviews by IANA Functions Operator and release request for implementation

This time, which reflects the “manual processing” component of the workflow, is the time when requests are pending review by staff, being actively reviewed by staff, or otherwise being processed in a manner in which PTI is responsible for the timeliness but is not part of the automated processes like contact confirmations and technical checking.

For routine requests, this component includes performing manual reviews, well formedness checks, responding to any requests for special exemptions, responding to requests for clarification from the customers, and the time associated with adhering to regulatory obligations.

**Important note:** Due to the methodology in which the measurement is performed, time during the preliminary period that is attributable to NTIA is being counted as staff processing time. Post-transition, this time will no longer be a factor which is expected to reduce the staff processing time. Therefore this time should be re-evaluated when there is sufficient post-transition data to consider.

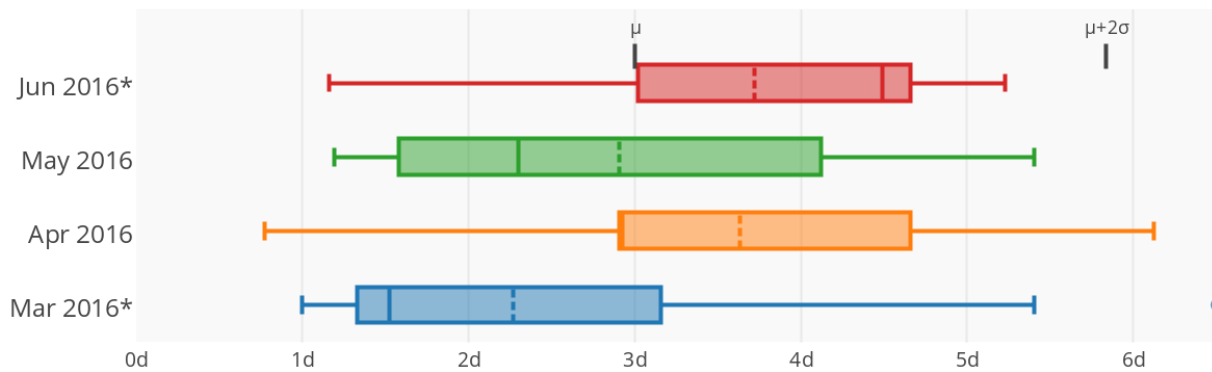
For Categories I through III, we propose performance thresholds based on the mean plus three standard deviations, albeit with a 90% requirement. This reflects the higher variability of this measure compared with other measurement types.

No ccTLD delegation or transfer requests were processed in the period, and therefore no conclusions can be drawn from the data. Therefore we have considered the existing performance standards under the NTIA IANA Functions Contract which have been considered satisfactory by the community, in which end-to-end performance of Category IV requests is expected to be performed within 120 days 50% of the time. We think it would be reasonable to set a starting benchmark at half this value for IANA processing time, until further data is available and the post-transition environment for ccTLD delegations and transfers is clarified.

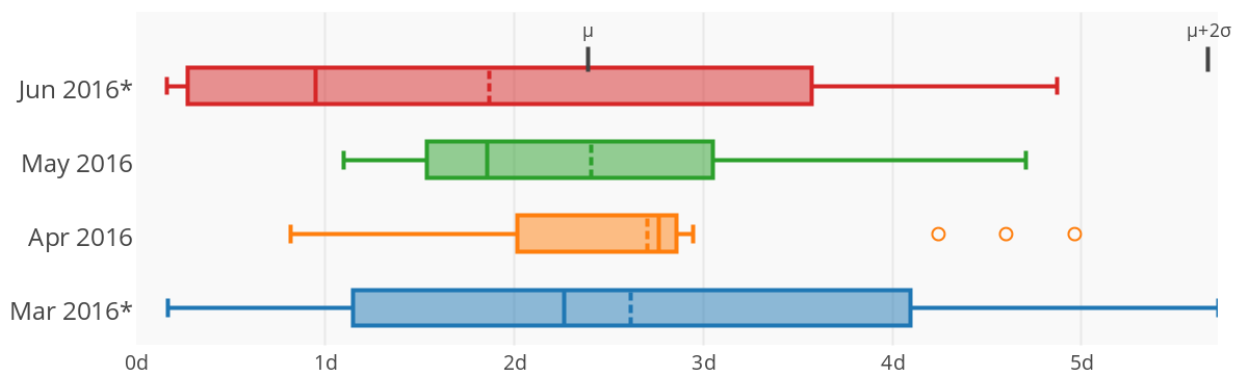
Category V, of which there was one request during the time period, is expressly intended to be a category for which there is unusual IANA processing for which a performance threshold would not be appropriate. Therefore none is recommended.

Category	Data	Recommendation
I	n=131, mean=3.00d, min=0.77d, max=6.51d, pstdev=1.42d, +2s=5.84d, +3s=7.26d	≤ 7d (90%)
II	n=81, mean=2.39d, min=0.16d, max=5.72d, pstdev=1.64d, +2s=5.67d, +3s=7.31d	≤ 7d (90%)
III	n=103, mean=4.37d, min=0.12d, max=17.13d, pstdev=2.57d, +2s=9.51d, +3s=12.08d	≤ 12d (90%)
IV	n=0	≤ 60d (50%)
V	n=1, mean/min/max=4.36d	None.

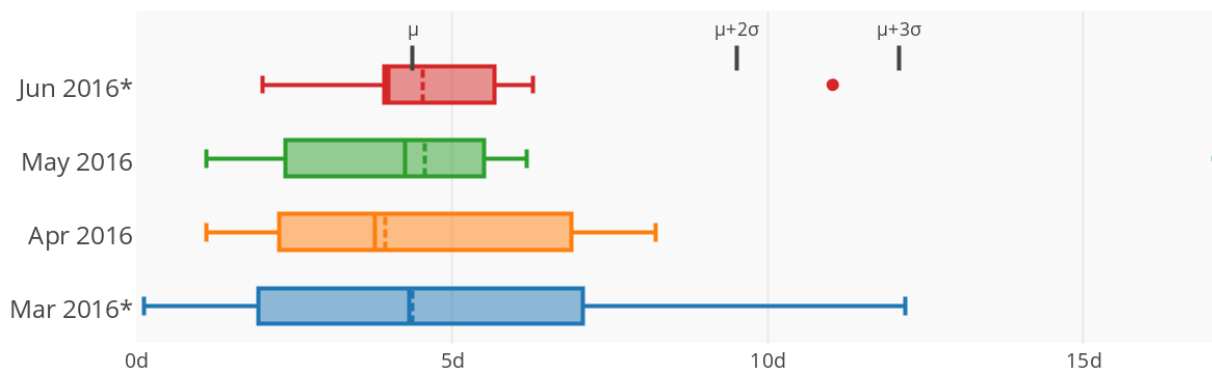
Validation and Reviews (Cat I)



Validation and Reviews (Cat II)



Validation and Reviews (Cat III)



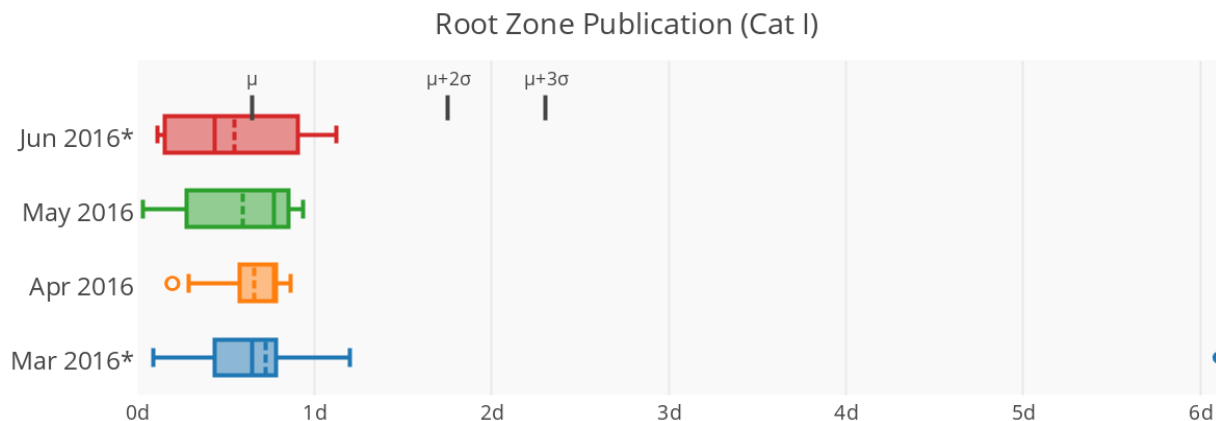
## Processing Performance (Implementation of Changes)

### Time for root zone changes to be published following completion of validations and reviews by IANA Functions Operator

The Root Zone Maintainer Agreement<sup>3</sup> between ICANN and Verisign stipulates performance of this component at  $\leq 72$  hours for 99% of performance. As the agreement wholly governs how quickly this process is performed, this threshold should be set to match this until superseded by changes to the agreement or further requirements setting by the community.

There was one significant outlier of over six days for an individual request. This relates to a change that was deliberately held for later implementation to align with another change request at request of the customer. There may be benefit post-transition in either formalizing deferred-implementation requests as a distinct category or considering them as Category V requests due to their infrequency.

Category	Data	Recommendation
I	n=131, mean=0.65d, min=0.03d, max=6.10d, pstdev=0.55d, +2s=1.75d, +3s=2.31d	$\leq 72$ h (99%)
II	Not applicable.	—
III	n=103, mean=18.62h, min=0.71h, max=44.70h, pstdev=7.55h, +2s=33.73h, +3s=41.29h	$\leq 72$ h (99%)
IV	n=0	"
V	n=1, mean/min/max=17.87h	"



<sup>3</sup> [https://www.icann.org/iana\\_imp\\_docs/63-root-zone-maintainer-agreement-v-1-0](https://www.icann.org/iana_imp_docs/63-root-zone-maintainer-agreement-v-1-0)

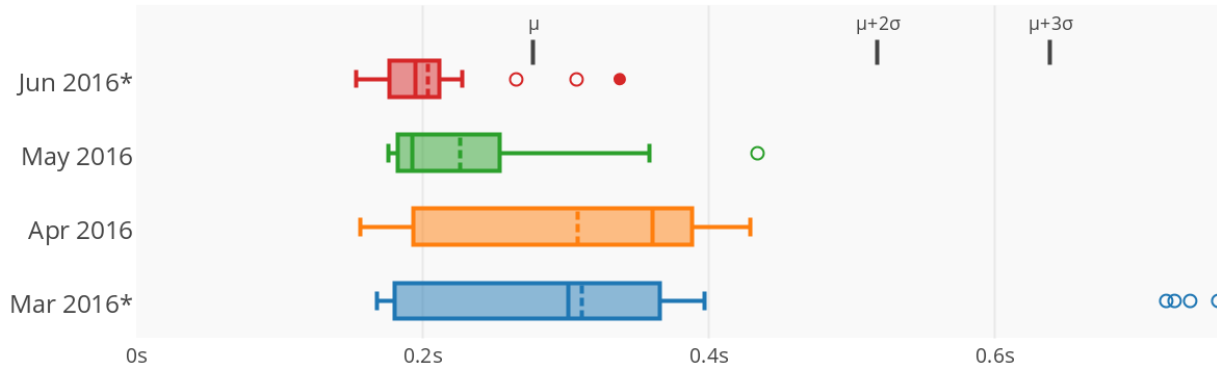


### Time to notify requester of change completion following publication of requested changes

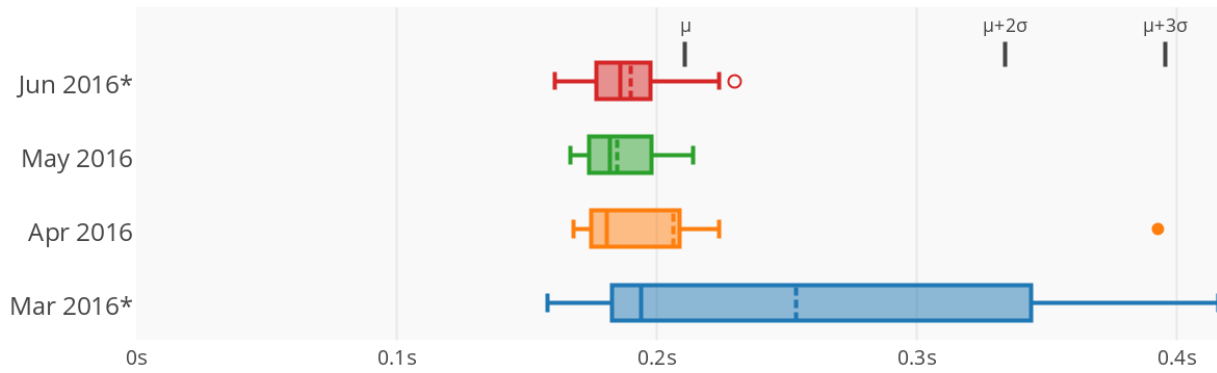
Measurements suggest this performance process typically takes 1-2 seconds to process. Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Category	Data	Recommendation
I	n=131, mean=0.28s, min=0.15s, max=0.76s, pstdev=0.12s, +2s=0.52s, +3s=0.64s	≤ 60s (95%)
II	n=81, mean=0.21s, min=0.16s, max=0.42s, pstdev=0.06s, +2s=0.33s, +3s=0.40s	"
III	n=103, mean=0.59s, min=0.42s, max=2.11s, pstdev=0.29s, +2s=1.17s, +3s=1.47s	"
IV	n=0	"
V	n=1, mean/min/max=0.45s	"

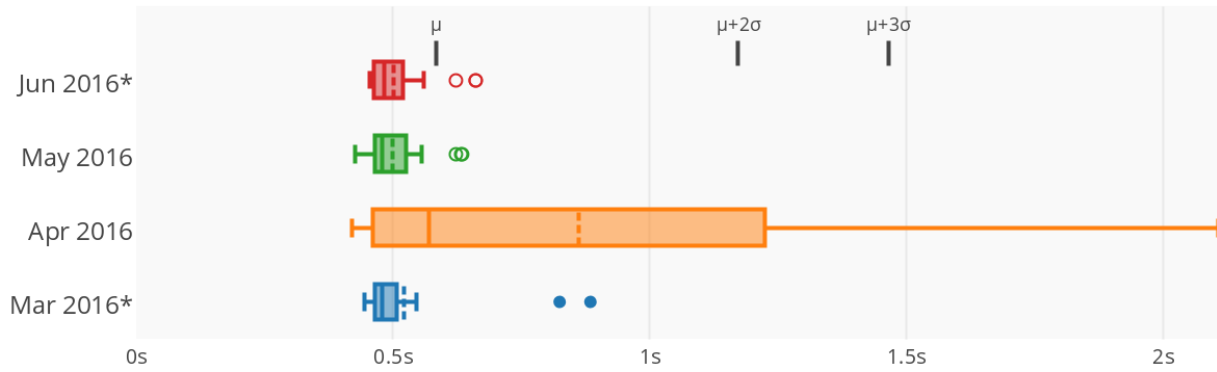
Notification of Completion (Cat I)



Notification of Completion (Cat II)



Notification of Completion (Cat III)





## Accuracy

### Root zone file data published in the root zone matches that provided in the change request

This measures that the technical data supplied by the customer, and validated throughout the root zone management process, is what is implemented in the root zone.

Category	Data	Recommendation
I	$\bar{x}=100\%$	= 100%
II	Not applicable.	—
III	$\bar{x}=100\%$	= 100%
IV	$\bar{x}=100\%$	"
V	$\bar{x}=100\%$	"

### Root zone database is correctly updated in accordance with change requests

This measures that the non-technical data supplied by the customer, and validated throughout the root zone management process, is what is stored in the root zone database (and subsequently recorded in the RZMS, WHOIS service, etc.)

Category	Data	Recommendation
I	$\bar{x}=100\%$	= 100%
II	$\bar{x}=100\%$	"
III	$\bar{x}=100\%$	"
IV	$\bar{x}=100\%$	"
V	$\bar{x}=100\%$	"

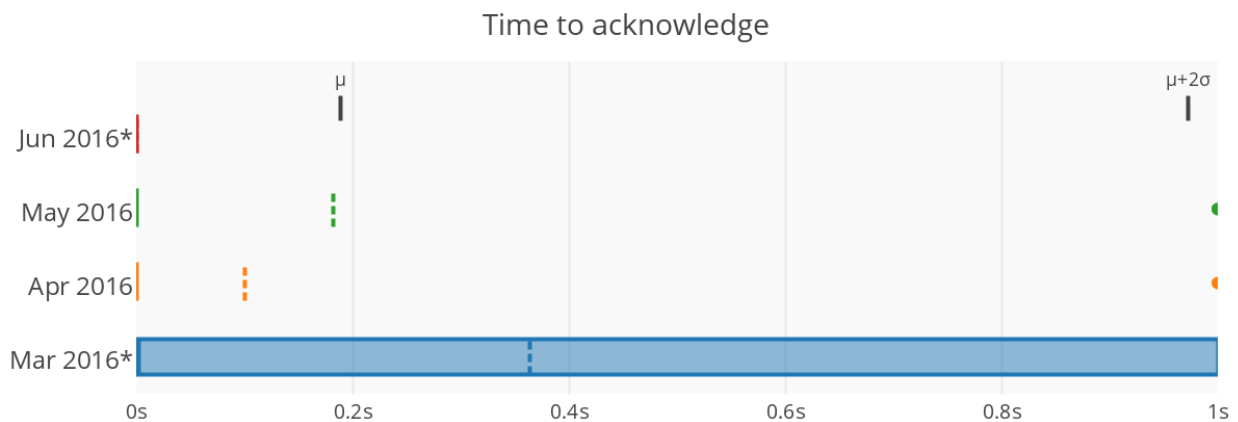
## Enquiry Processing

### Time to send acknowledgement of enquiry

This is a wholly automated process that typically takes less than a second. Variability is likely associated with minor fluctuations in system load, as enquiries are lodged in a shared system used across all of the IANA functions. Note that the ticketing system used for enquiry handling only measures the relevant events to the second, therefore we do not have sub-second granularity for this measurement category.

Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Data	Recommendation
n=37, mean=0.19s, min=0.00s, max=1.00s, pstdev=0.39s, +2s=0.97s, +3s=1.36s	≤ 60s (95%)

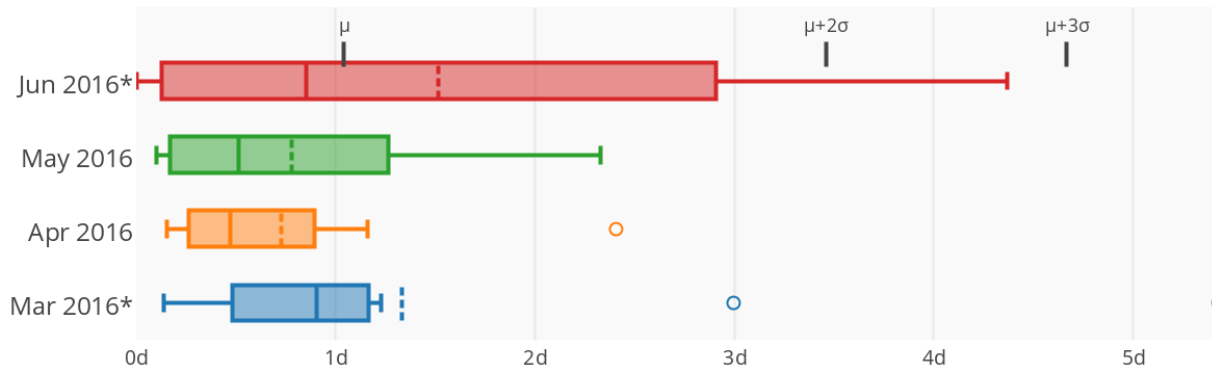


### Time to send initial response to enquiry

Typical response times average to 1-2 days, but can vary depending on whether an enquiry is received on a weekend and the complexity of the enquiry — some enquiries are escalated to internal subject matter experts in order to provide a more substantive initial response.

Data	Recommendation
n=33, mean=1.04d, min=0.00d, max=5.42d pstdev=1.21d, +2s=3.46d, +3s=4.67d	≤ 5d (90%)

# Enquiry Processing



## System Availability

System Availability is measured to measure all availability, and does not consider the difference between scheduled maintenance and unscheduled downtime. We have specified these at 99%, which reflects up to 7.2 hours of permissible downtime per month.

### RZMS availability

Data	Recommendation
March 2016: 99.96%	≥ 99.0%
April 2016: 99.96%	
May 2016: 99.97%	
June 2016 (partial): 100%	

### Website availability

Data	Recommendation
March 2016: 100%	≥ 99.0%
April 2016: 100%	
May 2016: 100%	
June 2016 (partial): 100%	

### Directory service availability

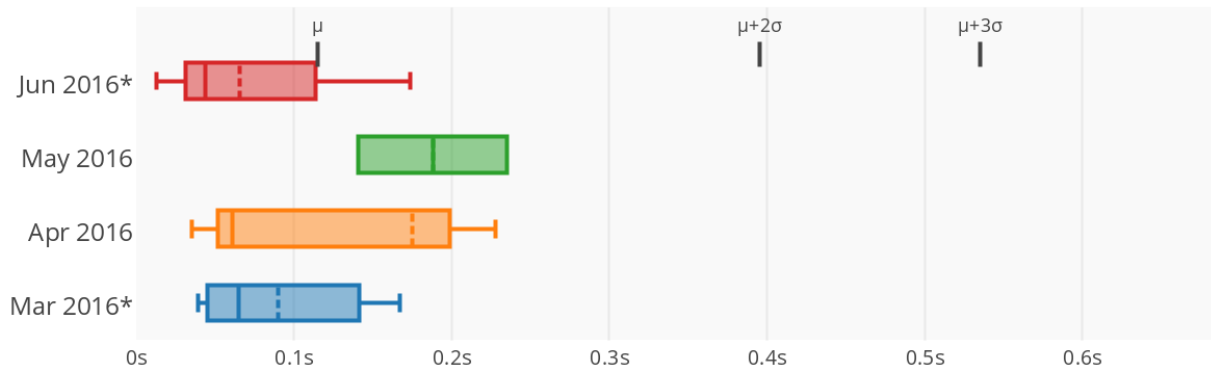
Data	Recommendation
March 2016: 100%	≥ 99.0%
April 2016: 100%	
May 2016: 100%	
June 2016 (partial): 100%	

### Credential recovery

The measure reflects “time to dispatch confirmation email of forgotten username and password”. This is a wholly automated process that took no longer than one second for all 22 invocations during the preliminary period. Given this is an automated process with little perceptible delay, we recommend an SLA of 60 seconds that would account for minor system variability that is not considered to impact customer experience.

Data	Recommendation
n=22, mean=0.12s, min=0.01s, max=0.69s, pstdev=0.14s, +2s=0.40s, +3s=0.54s	≤ 60s (95%)

## Credential Recovery



### Credential change

The measure reflects “time to implement new password within the system”, and is intended to represent the time between when the customer clicks “submit” on a web form with their new password until the time the password is inserted into the user database as a result. This is a wholly automated process and is anticipated to reliably also only take fractions of a second, however, the instrumentation of this event was identified to have not been implemented correctly and ICANN does not have reliable data at this time. We recommend this be re-evaluated once correct data is collected in a future version of RZMS.

Data	Recommendation
No data.	—

### SLE Dashboard Production

No data exists on update frequency, accuracy and availability as this activity will only occur post-transition.

### SLE Report Production

No data exists on update frequency and availability as this activity will only occur post-transition.