

ICANN's Work on Internationalising the Domain Name System

Kim Davies
Manager, Root Zone Services
ICANN



Internet Corporation for
Assigned Names & Numbers

What is ICANN?



What is ICANN?

- ▶ Created in 1998, headquartered in Los Angeles
- ▶ Multi-stakeholder organisation tasked with coordinating the allocation of Internet unique identifiers
- ▶ Runs the “Internet Assigned Numbers Authority” (IANA)
- ▶ Domain Names, IP addresses, Port Numbers, URI schemes, Private Enterprise Numbers, etc.



Domain Name System

- ▶ ICANN oversees assignment of top-level domains (.com, .org, .uk, .ca)
 - ▶ These are listed in the DNS “root zone”
- ▶ Involved in various other aspects, accreditation of domain registrars, overseeing policy for generic top-level domains, etc.

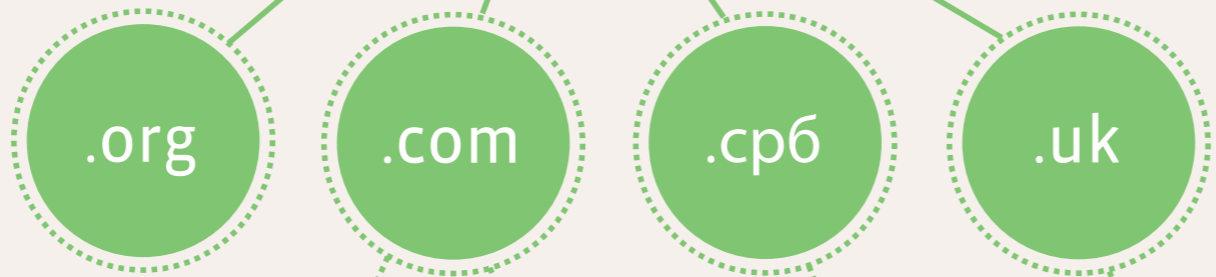


DNS Root



Managed by ICANN

Top Level Domains



Second Level Domains



Third Level Domains



Internationalising the DNS Root Zone



IDNA Protocol

- ▶ Domain Name system historically a subset of ASCII known as “LDH”.
- ▶ IDNA protocol overlays the standard DNS, providing an LDH-safe encoding of Unicode strings
- ▶ First version of IDNA protocol released in 2003, second major revision in 2010.



U-label

例え.テスト



xn--r8jz45g.xn--zckzah

A-label



IDNA Deployment

- ▶ Deployed at second level in registries (e.g. 日本語.jp) starting in 2003
 - ▶ However, this doesn't allow for fully internationalised domains in non-Latin script.
- ▶ Fully internationalised domains important to support intuitive addresses.
 - ▶ Don't need to mix scripts in a full domain name.
 - ▶ Exceptionally important for right-to-left scripts.



First test deployment in the root

- ▶ 11 translations of the word “test” were delegated in 2007
- ▶ 11 translations of <http://example.test> take you to a Wiki site for discussing test-related issues (see <http://idn.icann.org/>)



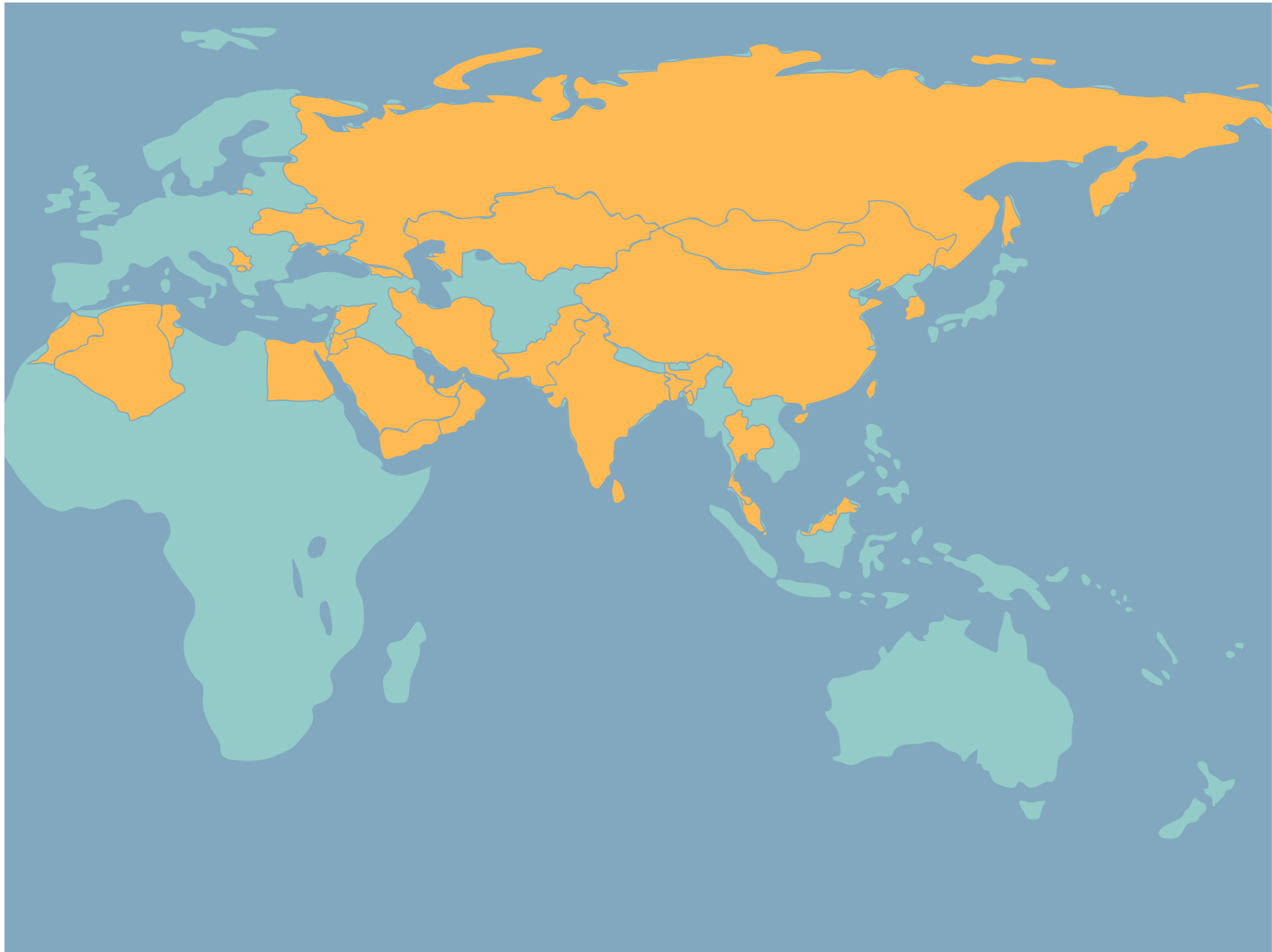
	Language	Script
http://مثال.إختبار	Arabic	Arabic
http://例子.测试	Chinese	Simplified Han
http://例子.測試	Chinese	Traditional Han
http://παράδειγμα.δοκιμή	Greek	Greek
http://उदाहरण.परीक्षा	Hindi	Devanagari
http://例え.テスト	Japanese	Katakana
http://실례.테스트	Korean	Hangul
http://مثال.آزمایشی	Persian	Arabic
http://пример.испытание	Russian	Cyrillic
http://உதாரணம்.பரிட்சை	Tamil	Tamil
http://בײַשפּיל.לעסען	Yiddish	Hebrew



“Fast Track” for country names

- ▶ Pressing need for production usage of domains was for representations of country names (e.g. .中国, .РФ) in native scripts.
- ▶ The formal policy development process in ICANN takes years.
- ▶ An interim “fast track” approach was developed to satisfy immediate need for country names in non-Latin scripts. First delegations occur in 2010.





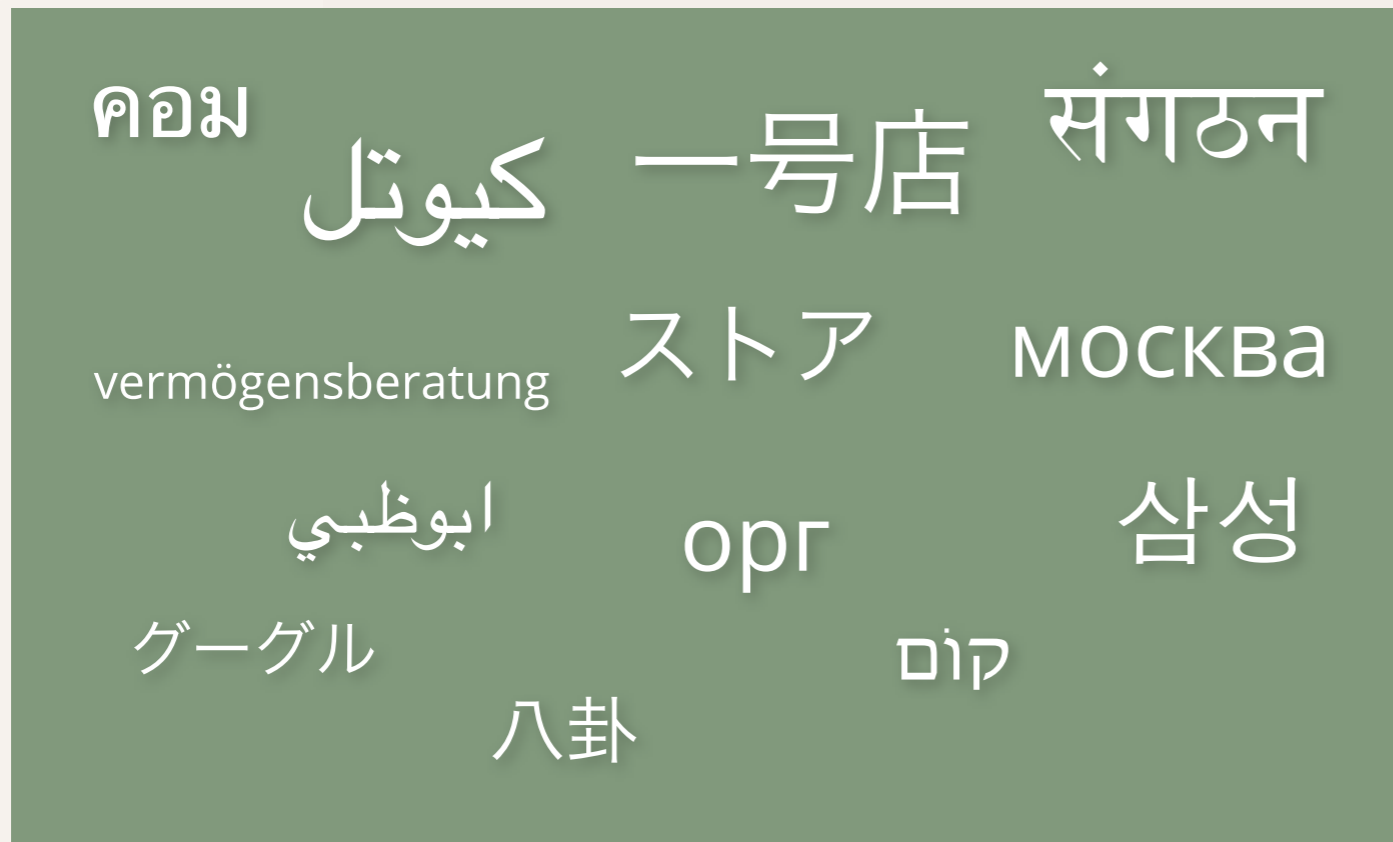
Fast Track has resulted in...

- ▶ 37 unique requests for country codes in non-Latin scripts
 - ▶ 47 strings approved
- ▶ 33 new country code top-level domains, representing 23 countries
- ▶ 23 languages represented, in 15 different scripts



New gTLD Programme

- ▶ In 2012, ICANN solicited applications for new gTLDs (“.anything”).
- ▶ Of 1930 applications, 116 are IDN strings

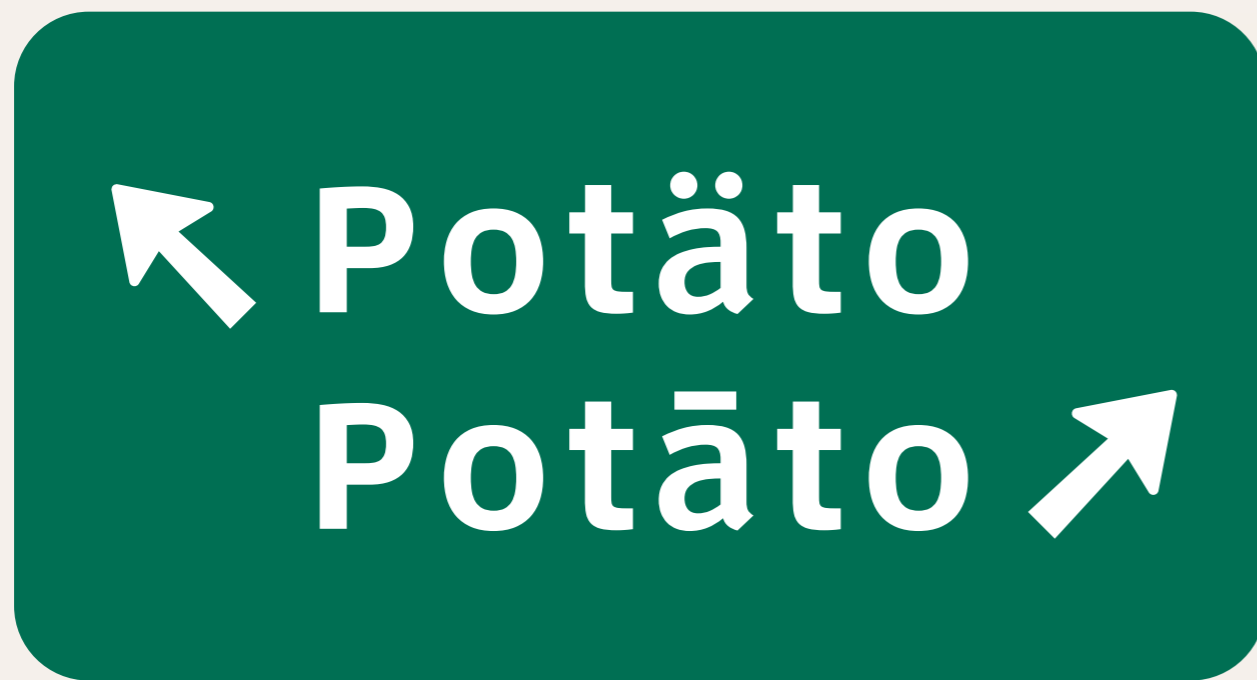


Current state of affairs

- ▶ “Fast track” country names deployed
 - ▶ Full domain names in native script available for several scripts today.
 - ▶ Long-term policy being developed
- ▶ Significant number of new generic TLDs pending evaluation



Where we hit difficulties...



Ə

U+01DD

Ə

U+0259



xn--8ja



xn--sna



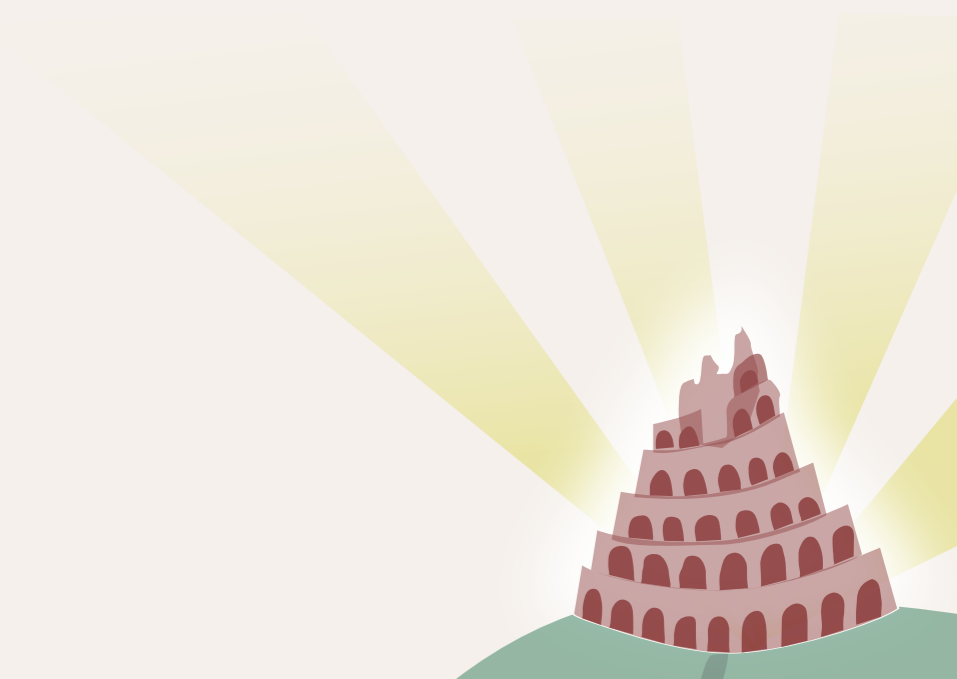
IDNA protocol operation

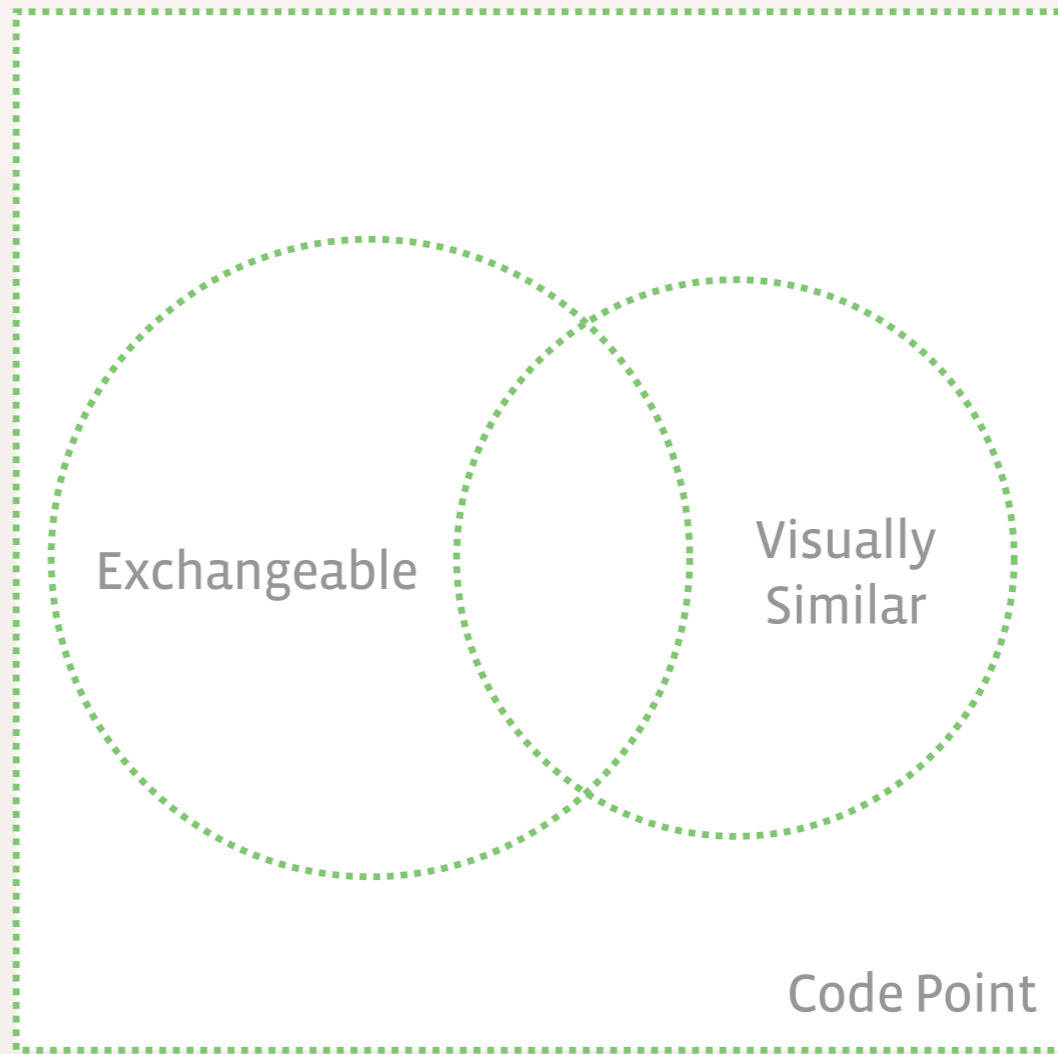
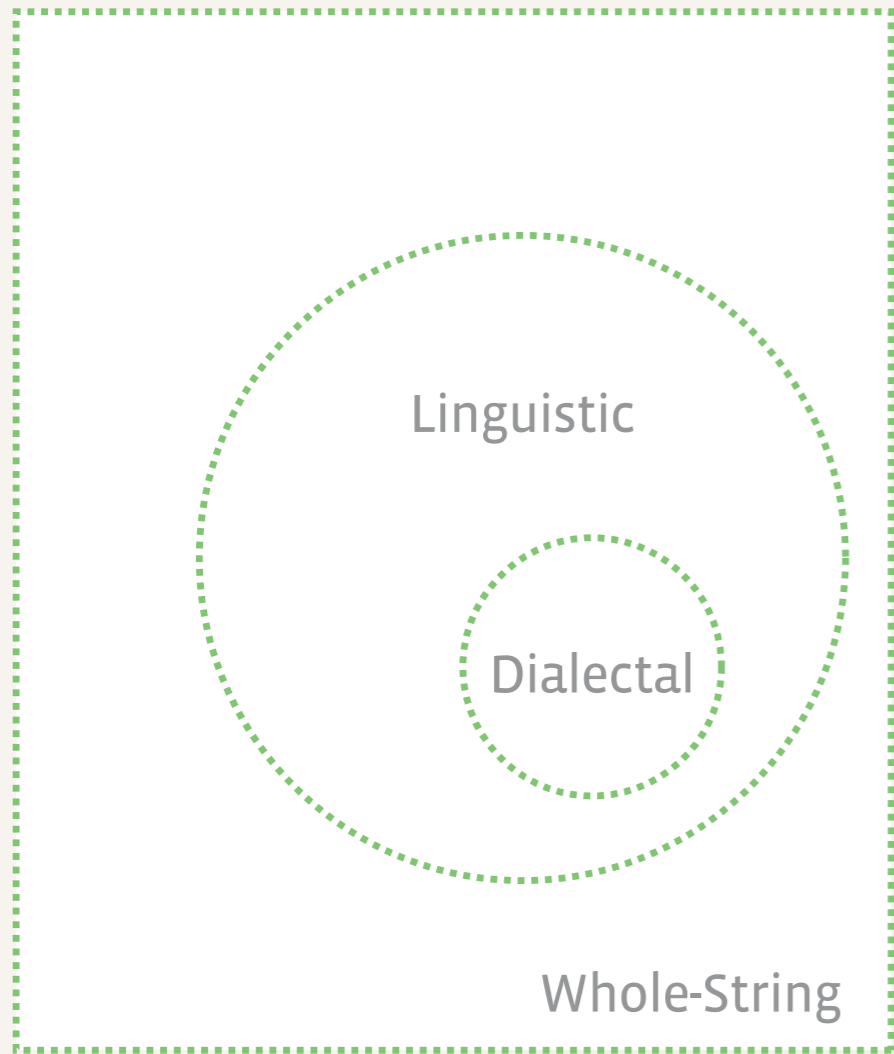
- ▶ Similar or visually identical strings can result in two or more distinct A-labels.
 - ▶ And, multiple methods of entering the “same” string can result in two or more distinct A-labels.
- ▶ DNS provides no mechanism to connect one domain (and all its children) to another, and enforce this down the delegation tree.
- ▶ The root zone is unique. Most domains are language specific, the root is a shared resource globally.

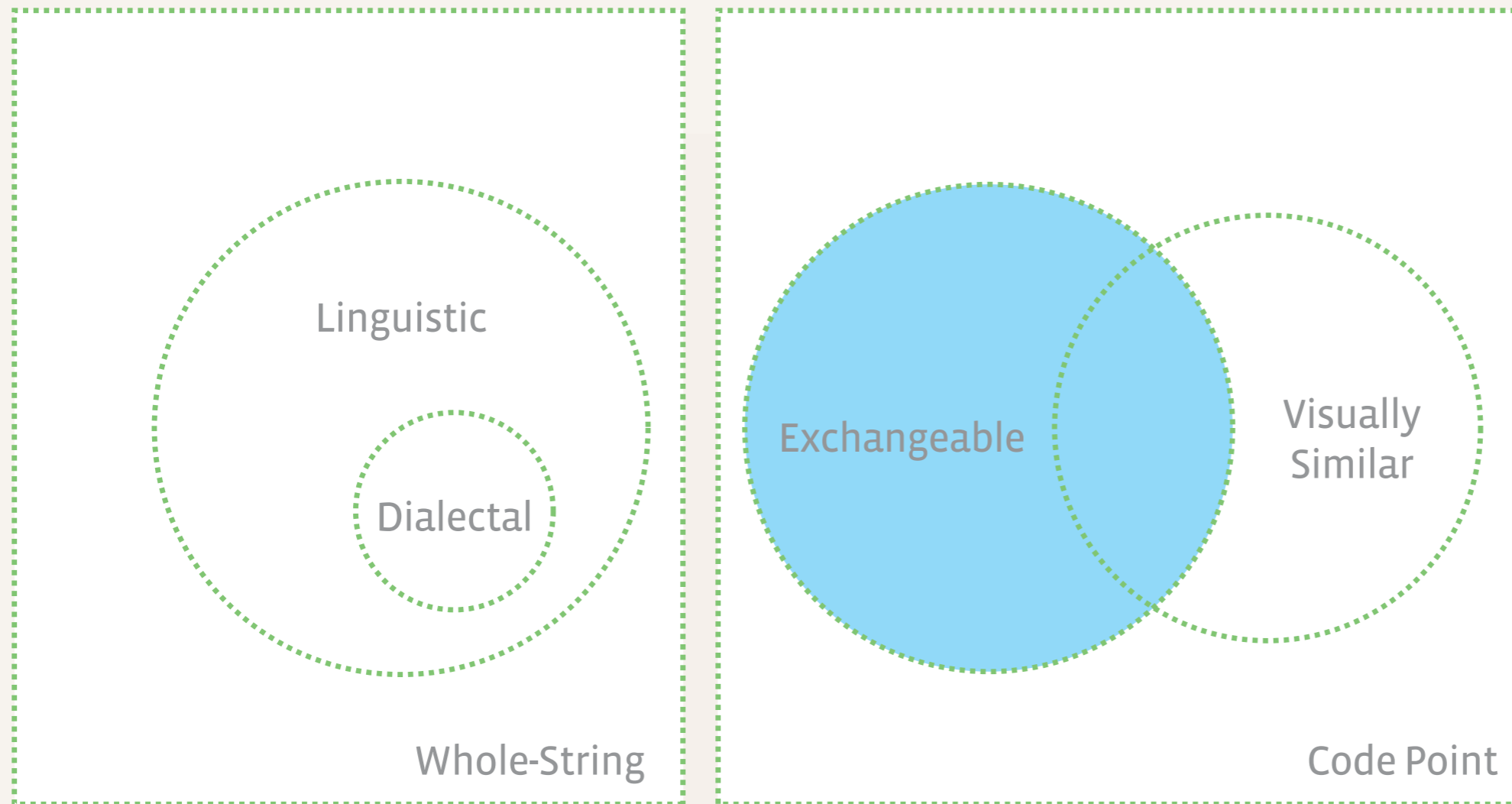


“Variants”

- ▶ The multiple different potential representations are variously known as “variants”
- ▶ A taxonomy of potential variants has been developed as part of our project work

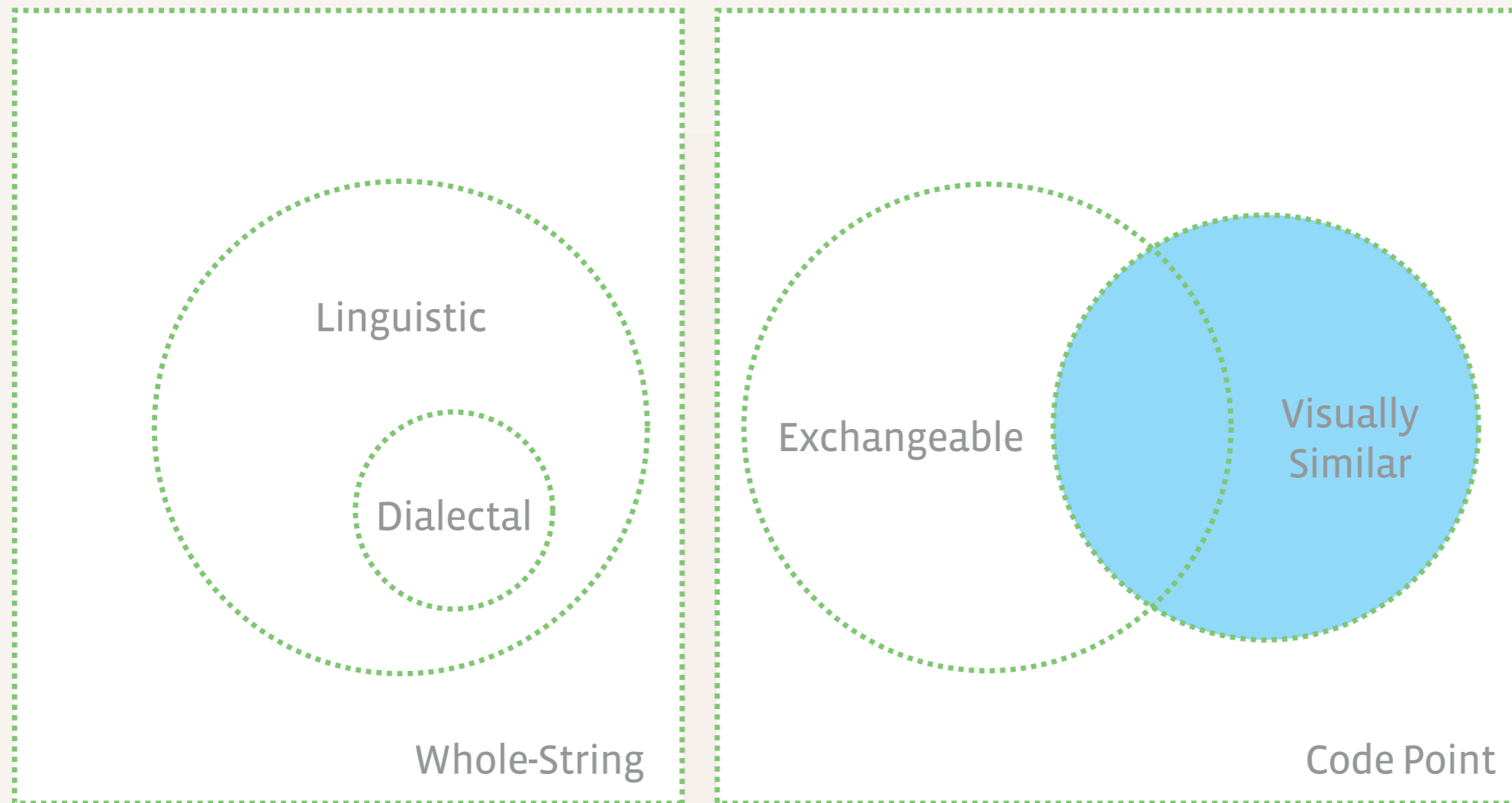






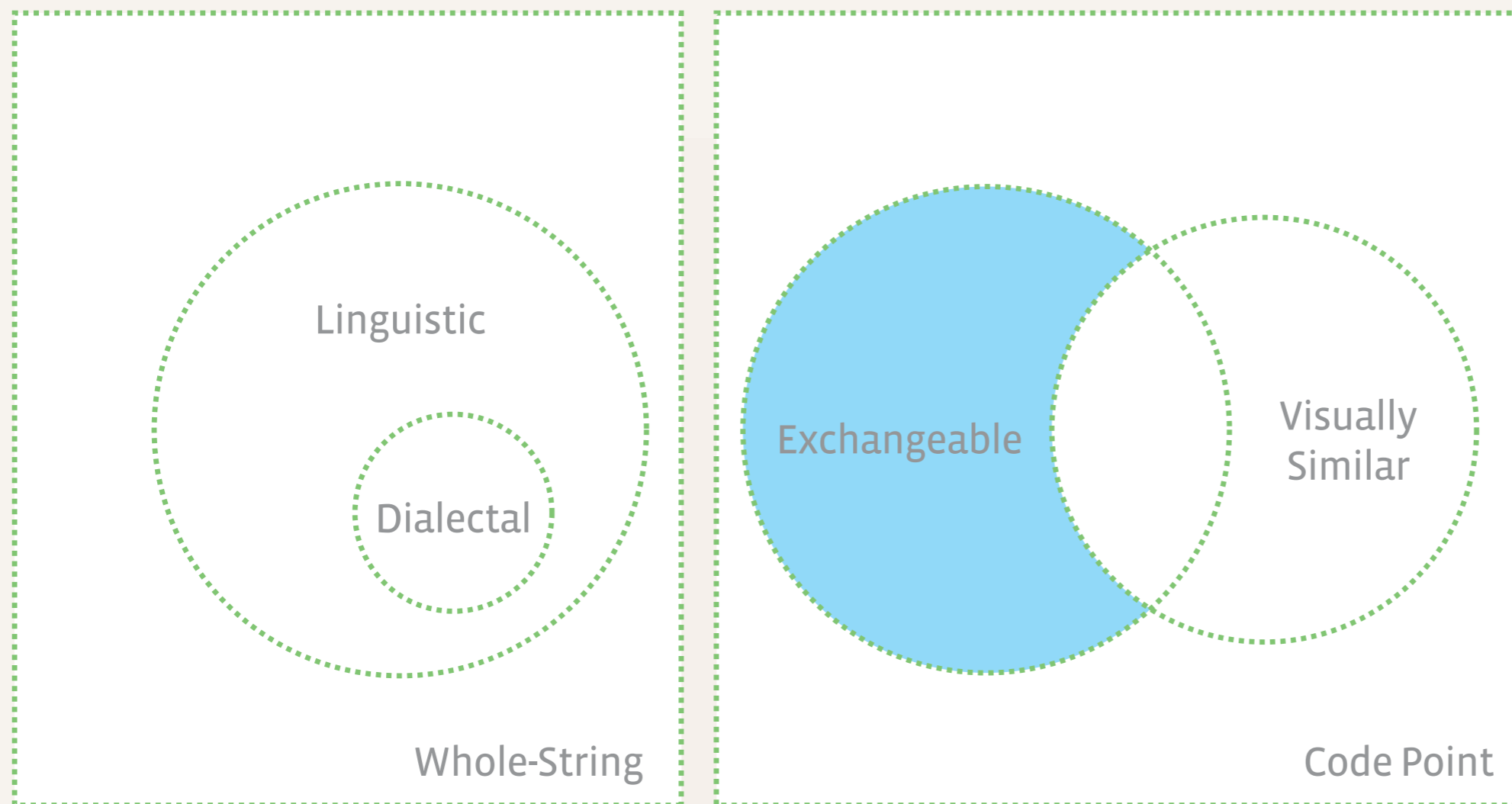
- ▶ **exchangeable** are two or more code points are seen by a user as so closely related that they may fill the same role in a domain name.





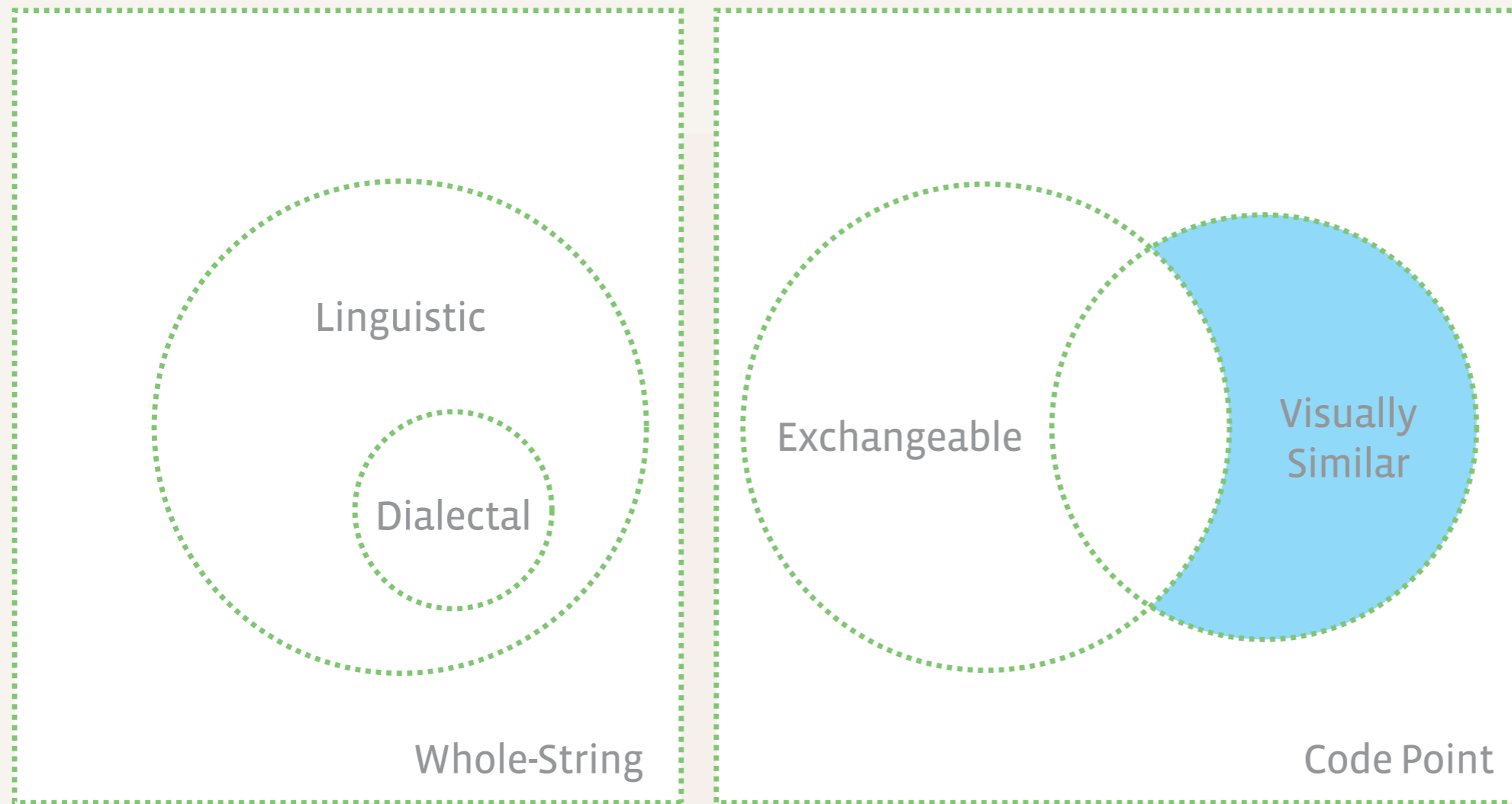
- ▶ **visually similar** are two or more glyphs are so much alike that they may be mistaken for one another.





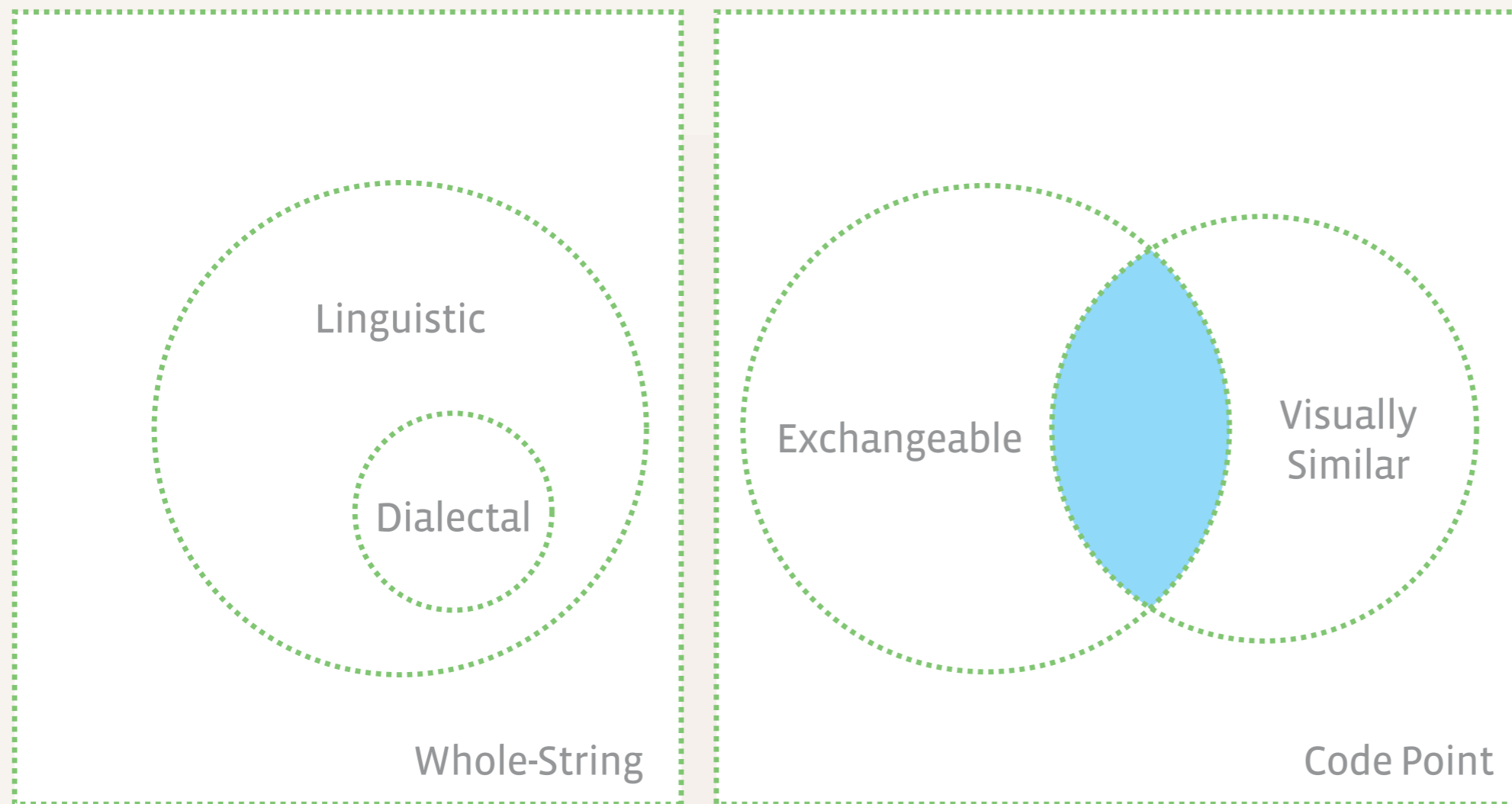
- ▶ **exchangeable but not visually similar**
- ▶ compatibility mappings, join-control characters, upper/lower case and underspecified information, positional variants (Greek sigma/final sigma)





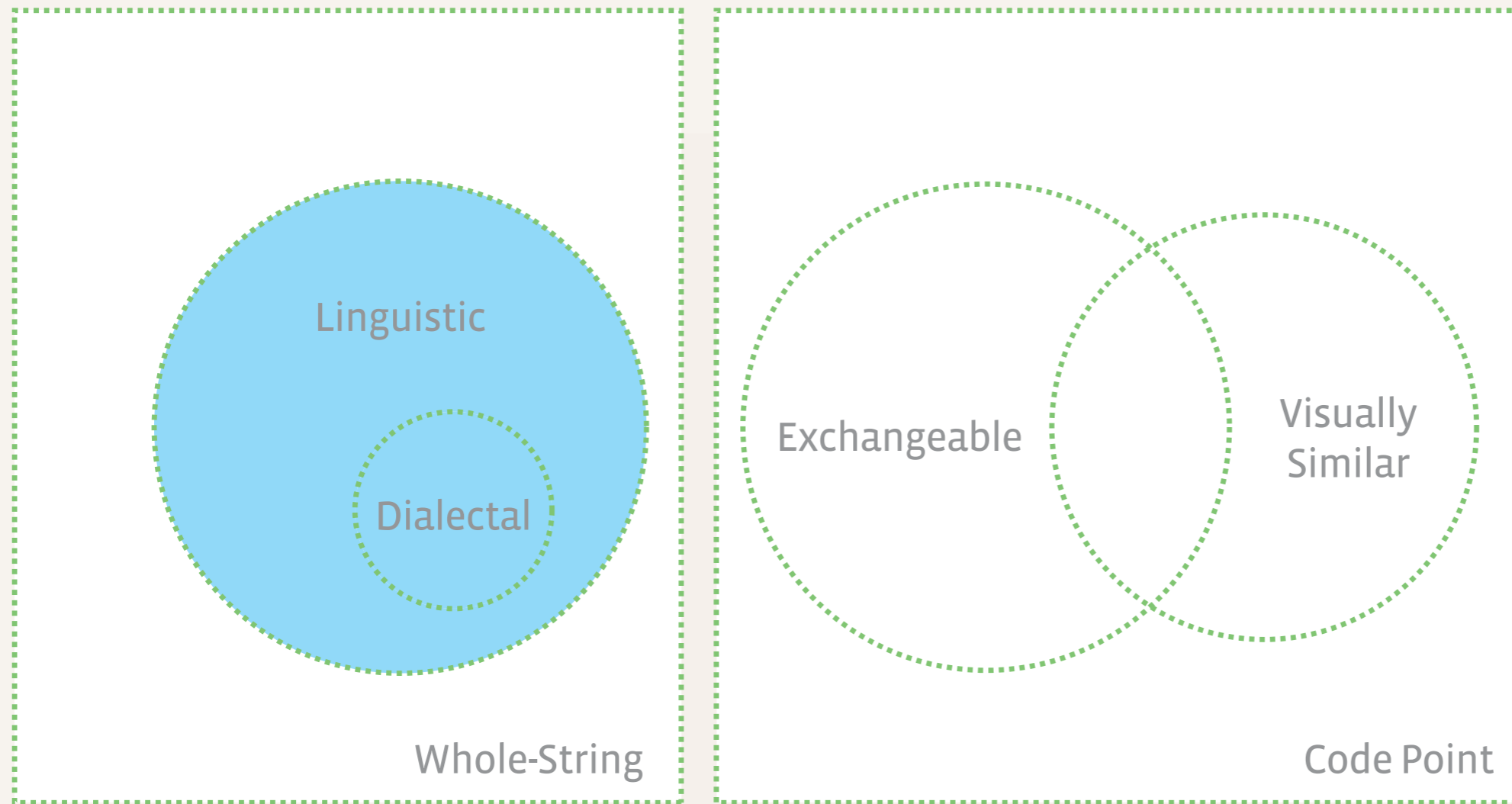
- ▶ **visually similar but not exchangeable**
- ▶ simple visual similarity, inter-script confusables between scripts (Greek, Latin, Cyrillic; Devanagari, Gujarati), linguistic variants (Dimotiki, Katharevousa)





- ▶ **exchangeable and visually similar**
- ▶ same abstract character w/ different encodings (turned e & schwa), same abstract character differently rendered in some contexts, different characters but exchangeable by users (hamza, alef, simplified/traditional Han, ghe & ghe with upturn), special characters (Ukr. apostrophe, Deva. apostrophe)





- ▶ **whole string variants** where variant token is larger than a character, i.e. morpheme, full word, phrase
- ▶ colour/color, Πειραιάς/Πειραιεύς, etc.



How could variant labels be treated?

- ▶ **Blocking** — not allowed to be used in any context
- ▶ **Withheld** — not used, but held only for registrant of the fundamental label
- ▶ **Allocated** — registered, but not used
- ▶ **Activated** — in the DNS
- ▶ **Delegated** — in the DNS as an NS record set
- ▶ **Mirrored** — method used to ensure corresponding trees are in sync



It's not just the DNS

- ▶ There is no perfect “variant” solution for the DNS protocol, and the problem has wider scope.
 - ▶ SMTP — configuring mail servers
 - ▶ HTTP — configuring web servers
 - ▶ End-user applications
 - ▶ etc. etc.



Current state of affairs

- ▶ Two sets of “variants” were delegated in the fast track, on the condition they be “synchronised”
 - ▶ This means the registry must keep the contents of the zone files with matching variants, and ensure that registrants below them follow this practice also.
 - ▶ Must be done contractually etc. as there is no technical mechanism to enforce this.
- ▶ Variants are not allowed* in current round of new gTLDs.



Future state of affairs

- ▶ The **Variant IDN Project** (VIP) was launched by ICANN to identify long term approaches.
- ▶ Firstly, a set of “issues” with variants was developed throughout 2011 until early 2012.
 - ▶ Studied six specific scripts (Devanagari, Latin, Cyrillic, Greek, Arabic, Han)
 - ▶ Issues were combined into an integrated issues report
- ▶ Now, this work has led to the creation of a number of new work activities to develop a long-term approach to supporting variant TLDs.



The Variant Projects



Label Generation Rules Process

- ▶ “Develop the process to define allowed code points, corresponding exchangeable variant code points, and related allowed states for IDN Variant TLDs”.
- ▶ Project convened in August
- ▶ First draft published in late September for community comment; second draft anticipated late November
- ▶ Final version expected in April 2013



Current draft process

- ▶ Overall process envisages two stage review:
 - ▶ **Primary panel** comprised of experts/representatives of a specific writing system
 - ▶ Job is to identify the requirements on behalf of the community that wishes to use the script
 - ▶ Can be divided into sub-panels, e.g. per language
 - ▶ **Secondary panel** has general expertise on entire Unicode, DNS, IDNA and Linguistics
 - ▶ Job is to review primary panel output, identifying impacts and practicality issues on the entire DNS system.



Study User Experience of Active Variants

- ▶ If two or more variants are delegated, what are the user experience implications?
 - ▶ End users
 - ▶ Power users (system admins, developers, etc.)
 - ▶ Registries, registrars, registrants
- ▶ What are the necessary rules or guidelines a TLD should operate under to provide an acceptable user experience?



User Experience Plans

- ▶ Two-phase report
 - ▶ Interim report focuses on usability principles and issues — October 2012 target
 - ▶ Draft final report focuses on recommendations — February 2013 target
- ▶ Each phase of the report will be published for public comment



Label Generation Tool

- ▶ IANA currently published “IDN tables” that list code points and variants for existing registries
 - ▶ Lacks a common format and standard implementation
- ▶ Proposing an XML-based standard for nominating code points and variants
- ▶ Unsure of publication path, just looking at creating something useful
- ▶ Ideally, will be the future IANA repository format
- ▶ May be an input into future root practice (e.g. by merging various tables)



The future of variant work



Concluding the variant work

- ▶ The active work on LGR Process, LGR tools, user experience will feed into an implementation approach for ICANN
- ▶ Some work has been ruled out of scope: Whole string variants, mirroring variants; due to lack of feasibility



Final thought on variants

- ▶ **meaningful mnemonics**, not representing the details of every language
- ▶ The goal of the DNS is **reliable matching**
- ▶ “Constraint and limitation is our friend in the domain name world”



Universal Acceptance



Universal Acceptance

- ▶ Delegating new domains and variants is not useful if software does not support it.
- ▶ Includes, but not limited to, supporting the IDNA protocol.
- ▶ Problem includes implementations using fixed TLD lists, TLD length checking, and other flawed approaches.
- ▶ How can ICANN best facilitate closing the gap to universal acceptance of all valid domains?



registry.sx - Google Search

www.google.com/search?aq=f&sourceid=chrome&ie=UTF-8&q=registry.sx

Did you mean to go to <http://registry.sx/>?

+You Search Images Maps Play YouTube News Gmail Documents Calendar More

Google registry.sx

Search About 5,220,000 results (0.26 seconds)

Web

SX Registry
registry.sx/
SX Domain Names .SX is the new autonomous country of Sint Maarten

FAQ
FAQ. General; Disputes; WHOIS; Trades & Transfers; Deleting

SX Registration Policies
Policies means these .SX Domain Name Registration Policies

[More results from registry.sx](#)

Marina Del Rey, CA
Change location

Show search tools

<http://example.com>

<http://example.xxx>

<http://япон.рф>

<http://example.XN--P1AI>



Permanent Home Address Line 1*:

Permanent Home Address Line 2:

City*:

State*: CA

Zip Code*: -

Note: You must be a U.S. resident. No P.O.

SSN*: - -

Date of Birth*: / / (MM/DD/YYYY)
You must be 18 or older, 19 in Al be cons

Mother's Maiden Name*:

E-mail Address: @ .

aero
 biz
 com
 coop
 edu
 gov
 info
 mil
 museum
 name
 net
 org
 pro
 us

But you can do much more with regular expressions. In a text editor like [EditPad Pro](#) or a specialized text processing tool like [PowerGREP](#), you could use the regular expression `\b[A-Z0-9._%+-]+@[A-Z0-9.-]+\.[A-Z]{2,4}\b` to search for an email address. Any email address, to be

All TLDs must be 2-4 ASCII letters (A-Z)



```
472 "tz", // Tanzania
473 "ua", // Ukraine
474 "ug", // Uganda
475 "uk", // United Kingdom
476 "um", // United States Minor Outlying Islands
477 "us", // United States of America
478 "uy", // Uruguay
479 "uz", // Uzbekistan
480 "va", // Vatican City State
481 "vc", // Saint Vincent and the Grenadines
482 "ve", // Venezuela
483 "vg", // British Virgin Islands
484 "vi", // U.S. Virgin Islands
485 "vn", // Vietnam
486 "vu", // Vanuatu
487 "wf", // Wallis and Futuna
488 "ws", // Samoa (formerly Western Samoa)
489 "ye", // Yemen
490 "yt", // Mayotte
491 "yu", // Serbia and Montenegro
492 "za", // South Africa
```

```
$_cctld = array(
    'ac',
    'ad','ae','af','ag',
    'ai','al','am','an',
    'ao','aq','ar','as',
    'at','au','aw','ax',
    'az','ba','bb','bd',
    'be','bf','bg','bh',
    'bi','bj','bm','bn',
    'bo','br','bs','bt',
    'bu','bv','bw','by',
    'bz','ca','cc','cd',
    'cf','cg','ch','ci',
    'ck','cl','cm','cn',
    'co','cr','cs','cu',
    'cv','cy','cz',
    'de','dk','dm',
```

Czechoslovakia?

```
$_gtld = array(
    'aero',
    'biz',
    'cat',
    'com',
    'coop',
    'edu',
    'gov',
    'info',
    'int',
    'jobs',
    'mil',
    'mobi',
    'museum',
    'name',
    'net',
    'org',
    'pro',
    'travel',
    'asia',
    'post',
    'tel',
    'geo',
)
```

.geo?



So how does one check valid domains?

- ▶ Do you need to check domain validity?
 - ▶ If not, don't do it. Rethink why you are checking.
 - ▶ Other aspects, such as email confirmation, will catch invalid domains.
- ▶ If so, is it an online application?
 - ▶ Online applications should check the DNS, which is always up-to-date and accurate.
- ▶ If a fixed list is needed, use a sustainable approach.
 - ▶ Root zone changes daily



Not to forget IDNs

- ▶ Multiple representations (U-label, A-label) of the same domain
- ▶ Not just in the “domain” field, but they can appear in email addresses, web addresses, name servers, in-line text.



Work so far

- ▶ ICANN has done some work on this so far, including basic software toolkits
- ▶ Workshops held to explain ICANN's historical work, and foster dialogue on where ICANN can best direct future efforts
- ▶ Input received will advise future work



Can you help?

- ▶ ICANN community is very focused on DNS provisioning, not good outreach to software vendors and other implementor communities
- ▶ What can ICANN learn from the Unicode community on how you have spread awareness of internationalisation support?
- ▶ What tools can ICANN create to best improve acceptance?

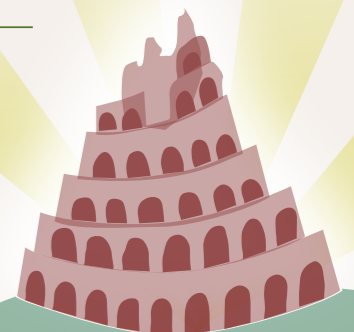


Resources



Variant Resources

- ▶ Integrated issues report on Variant IDN issues
 - ▶ <http://www.icann.org/en/news/announcements/announcement-20feb12-en.htm>
- ▶ Project Plan and Update
 - ▶ <http://www.icann.org/en/news/announcements/announcement-23aug12-en.htm>
- ▶ Draft LGR Process
 - ▶ <http://www.icann.org/en/news/public-comment/lgr-procedure-24sep12-en.htm>



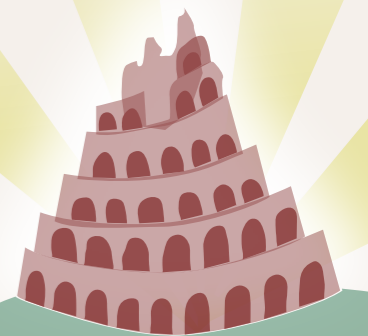
Label Generation Tool

- ▶ Current draft
 - ▶ <https://datatracker.ietf.org/doc/draft-davies-idntables/>
- ▶ Prototype implementation including sample converted tables
 - ▶ <https://github.com/kjd/idntables>



Universal Acceptance Resources

- ▶ Resource Centre
 - ▶ <http://www.icann.org/en/resources/tld-acceptance>
- ▶ Draft toolkits
 - ▶ <http://www.github.com/icann>



Thanks!

kim.davies@icann.org
@kjd



Internet Corporation for
Assigned Names & Numbers