OASIS CTI I18N

September 7th, 2016

Updated October 2016

i18n Proposals Summary

- Links
  - *i18n Specification - Pre-Draft* https://docs.google.com/document/d/1gNHSCGE3k6IhCjlYXFce3AfvkrabebmbulaipSr3_VY/edit?pref=2&pli=1#heading=h.9xt7wqfwufn9
  - *Thoughts on translations (John Wunder)* https://gist.github.com/johnwunder/6ded11efa5724826e8bcf269ee2f2edd
- Slack Channel - #i18n
- John-Mark's concerns on implementation (personally communicated)

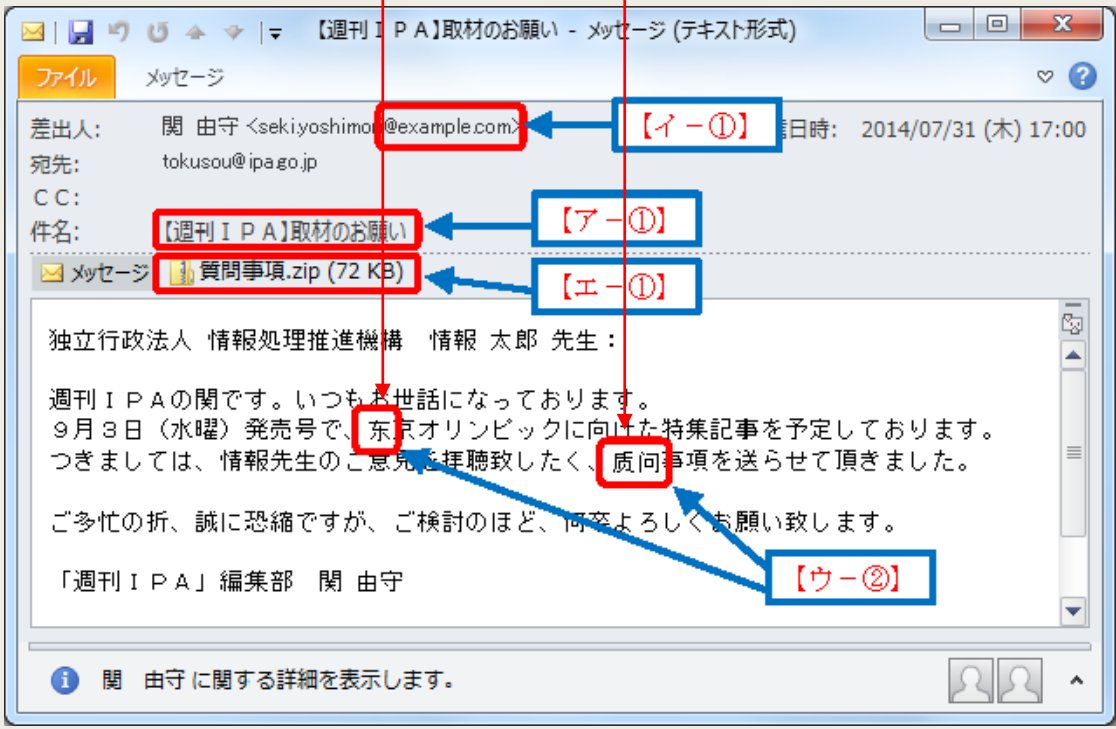
Source of ambiguity

- Han Unification of Unicode for CJK languages (*)
  - *“手紙” – “Letter” in Japanese, “toilet paper” in Chinese*
  - *Attacker use wrong characters in phishing emails*
    - 東京 → 东京, 質問 → 质问
  - The same group written differently
    - “Chinese Honker League” is “中国红客联盟” (cn), “中国紅客連盟” (ja), “中國紅客聯盟” (zh)
- “Hasta la vista” – All in ASCII, but in Spanish. People will know it, but it is not immediately apparent for computers

(*) https://en.wikipedia.org/wiki/Han_unification

Mistakes by Attackers

- 東京オリンピック → 东京オリンピック (Tokyo Olympic)
- 質問事項を → 质问事项を (queries)



Phishing Email Example

Reading Japanese

Alphabet | Katakana | Hiragana | Chinese Characters

techon.nikkeibp.co.jp/atcl/news/16/090203870/?n_cid=nbptec_tecml&rt=nocnt&d=147314160015

メガソーラー | FACTORY | SENSING | スポーツ | 5G | セミナー・技術者塾 | 協賛企業トピック

HOME > クルマ > 新車レポート > Aston Martin社「DB11」、サイドのラインは飾りにあらず

[Article List] 記事一覧 [New Car] [Report] 新車レポート

ニュース [News]

[Corp.] [Side] [Line] [frills]

Aston Martin社「DB11」、サイドのラインは飾りにあらず

[Reporter's Name] 窪野 薫 2016/09/02 17:12 1/1ページ [Page]

シェア 0 ツイート 保存

この記事どう？

5 ためになった

1 仕事に役立つ

知っておくべき

検索する

コメント投稿

印刷

その他

英Aston Martin社の日本法人は2016年8月31日、都内で新型スポーツカー「DB11」を発表した（図1）。従来車から空力性能を高めたことで、速度を上げても車体がぶれず、安定した走りができる（図2）。日本での価格は2380万円からを予定する。2016年8月25日にホンダが発表した、新型「NSX」の2370万円と同じ価格帯だ（**関連記事1**）。

図1 Aston Martin社の新型スポーツカー「DB11」

[画像のクリックで拡大表示]

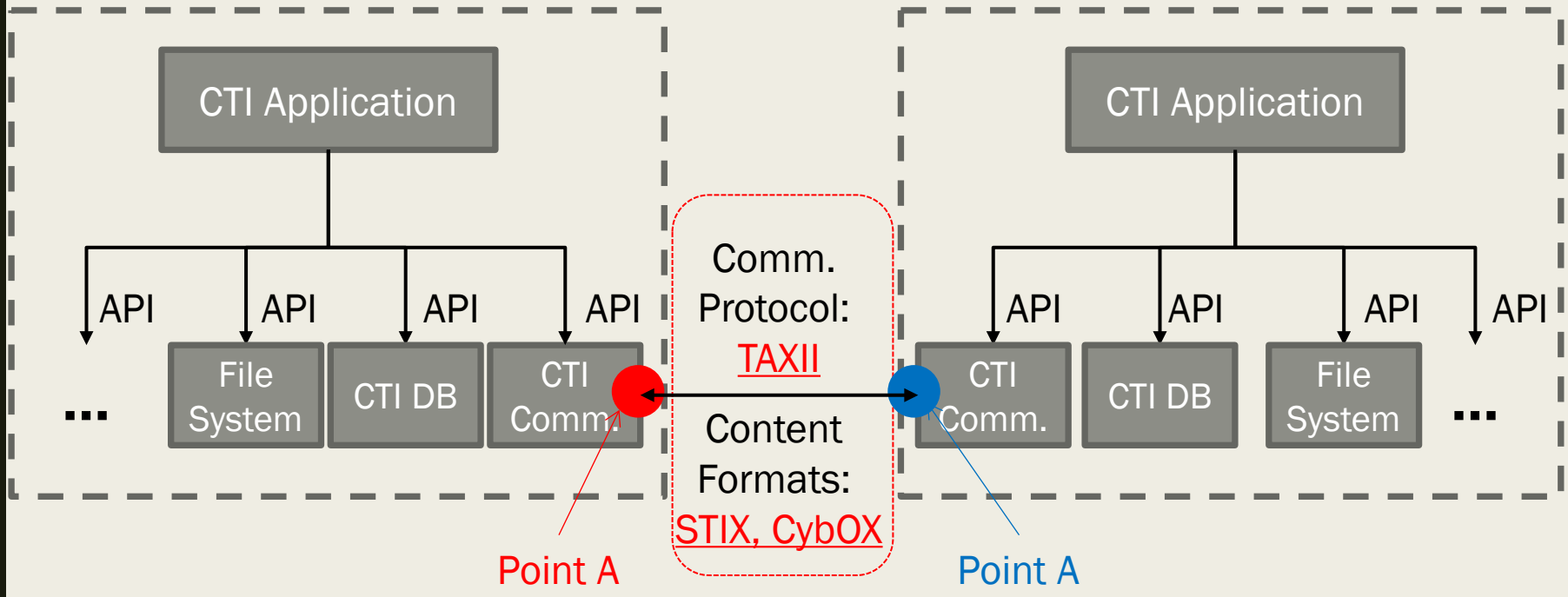
発表会に登壇した同社日本法人マネージングディレクターの寺嶋正一氏は、「DBシリーズの伝統的なサイドストレークと呼ぶラインは、従来車ではシンボルでしかなかった。DB11では飾りではなく、技術的な機能を持

おすすめ情報

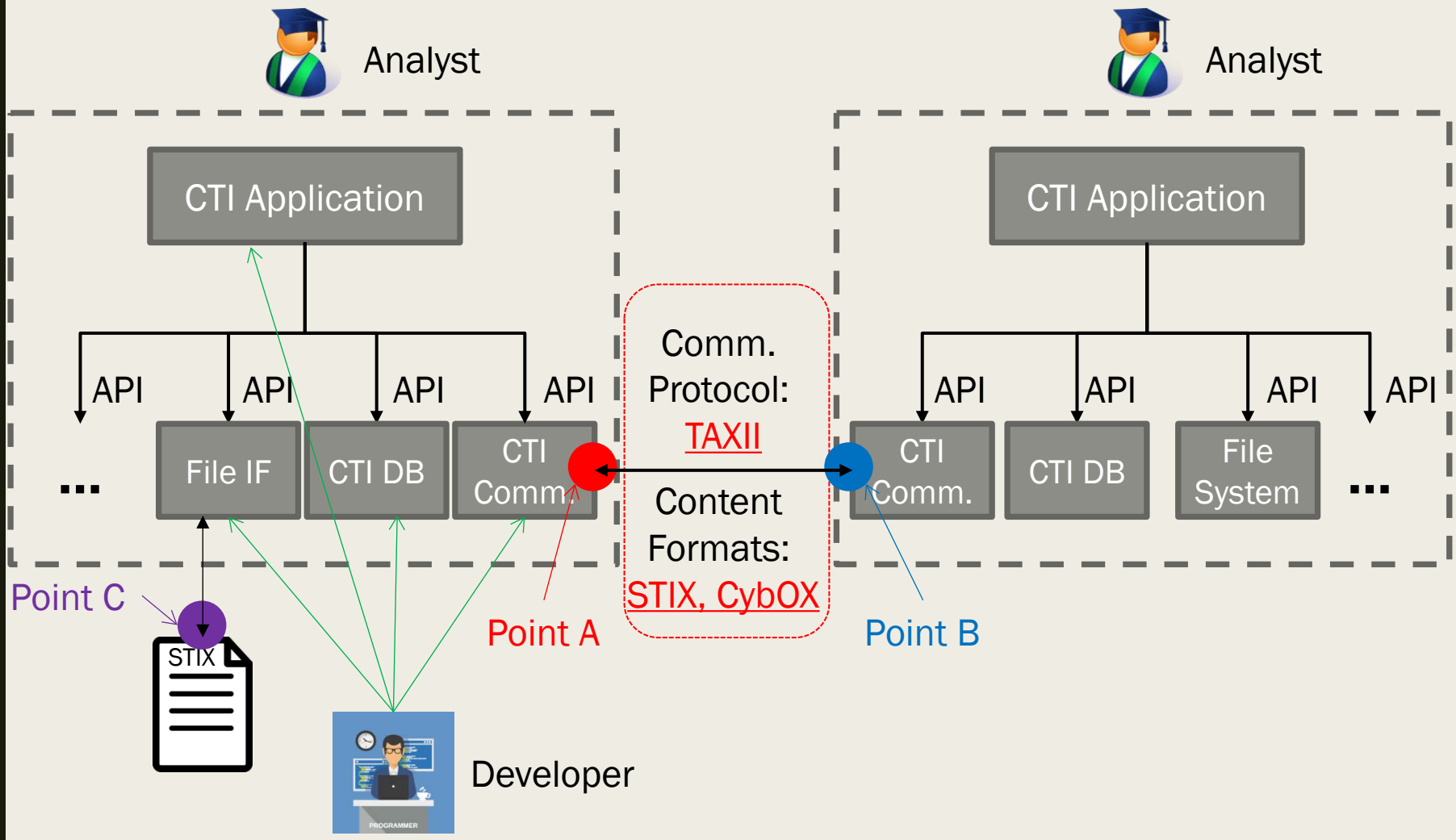
SCOPE

i18n Scope

- **Purpose:** CTI with "strings", including translations, communicated end-to-end semantically accurately
  - *TAXII, STIX, CybOX*
  - *APIs (CTI Comm., CTI DB, File system, etc.)*
  - *Parsing/generating CTI by CTI Applications*



Should Be Easy to Implement for Developers



Where in TAXII, STIX, CybOX - i18n Scope

- TAXII: Send and receive STIX as given binary data exactly to/from the other end
  - *Should be no i18n concern*
- CybOX: Whatever given between “ and “ are data, not a string to be interpreted (What to do with escaping?)
  - *Should be no i18n concern*
- STIX: There are several places where “strings” are used

Where in STIX – Common Data Types and Common Properties

Required in bold, *optional* in italic, ist of type string underlined

Common Data Type	Property	Section	Note
External Reference	**source_name**	2.3.1	"Fujitsu" or "富士通"
External Reference	*description*	2.3.1	Maybe different from the source_name language
External Reference	*url*	2.3.1	URL syntax
External Reference	*external_id*	2.3.1	In ASCII or identifier?
Kill Chain Phase	**kill_chain_name**	2.5	Ex. "高度標的型攻撃の攻撃シナリオ" (*)
Kill Chain Phase	**phase_name**	2.5	Ex. "攻撃立案", "攻撃準備", … (*)
Common Property	**type**	3.1	Must be a STIX Object type
Common Property	*labels*	3.1	Should come from a suggested vocabulary with additional labels

(*) https://www.ipa.go.jp/security/vuln/newattack.html

- 2.11 Open Vocabulary
  - *Do we allow languages other than English for open vocabulary (properties, values)?*
  - *Ex: Property: "攻撃-ラベル", Value: "標的型攻撃"*

Where in STIX – STIX Domain Objects (1/3)

Required in bold, *optional* in italic, ist of type string underlined

STIX Object	Property	Section	Note
Attack Pattern	**name**	5.1.1	
Attack Pattern	*description*	5.1.1	
Campaign	**name**	5.2.1	Name maybe in English (ex. “Blue Termite” or “NewsRipper”) …
Campaign	*description*	5.2.1	but, description (ex. “日本年金機構を含む…”)
Campaign	*objective*	5.2.1	and/or objective in Japanese
Course of Action	**name**	5.3.1	
Course of Action	*description*	5.3.1	
Identity	**name**	5.4.1	“Fujitsu” or “富士通”
Identity	*Labels*	5.4.1	e.g., CEO, Doctors, Hospital, or Retailer)
Identity	*description*	5.4.1	
Identity	*contact_information*	5.4.1	“担当者: 佐藤 03-1234-5678” (Sorry, I only speak Japanese)

Where in STIX – STIX Domain Objects (2/3)

STIX Object	Property	Section	Note
Indicator	**name**	5.5.1	
Indicator	*description*	5.5.1	
Indicator	**name**	5.5.1	
Indicator	*pattern_lang_version*	5.5.1	Ex. “1.0” for CybOX patterning language
Indicator	*pattern*	5.5.1	In a patterning language
Intrusion Set	**name**	5.6.1	
Intrusion Set	*description*	5.6.1	
Intrusion Set	*aliases*	5.6.1	What if it had a Chinese name?
Intrusion Set	*goals*	5.6.1	
Intrusion Set	*region*	5.6.1	“遼寧省” – Chinese region name (辽宁省) in ja?
Intrusion Set	*country*	5.6.1	Ex. “us”, “jp” from the ISO 3166-1 Alpha-2 codes
Malware	**name**	5.7.1	
Malware	*description*	5.7.1	

Where in STIX – STIX Domain Objects (3/3)

STIX Object	Property	Section	Note
Report	**name**	5.9.1	
Report	*description*	5.9.1	
Threat Actor	**name**	5.10.1	Ex. “中国红客联盟” as in original
Threat Actor	*description*	5.10.1	In English or Japanese
Threat Actor	*aliases*	5.10.1	Ex. “中国紅客連盟” (ja) and/or “Chinese Honker League” (en)
Threat Actor	*roles*	5.10.1	In English or Japanese
Threat Actor	*goals*	5.10.1	In English or Japanese (List of type string)
Tool	**name**	5.11.1	
Tool	*description*	5.11.1	
Tool	*tool_version*	5.11.1	Not much of an issue
Vulnerability	**name**	5.12.1	
Intrusion Set	*description*	5.12.1	

- Are they all? - We will surely have more properties in future

USE CASES

CTI i18n Use Cases

Multiple language property objects (the most contentious)

1. Providing string properties in multiple languages simultaneously at the time of creation
2. name and description, for example, in different languages

3. EN CTI received by a Japanese entity, which provides JA translation

4. An English CTI report describing attacks against Japanese entities in EN
5. Email subject/body, supposed to be in JP, but includes CN characters (by mistake of the attackers)

6. CTI translation service
7. CTI provider going global

Translation given later, separately, and/or on demand

8. Single-language producer

Case for keeping things simple

Single property value having multiple language elements

REQUIREMENTS

CTI i18n Requirements

- CTI Core Requirements
  - *Only one way to do things*
- i18n Specific Requirements
  - *Every message is in Unicode*
  - *Translations survive as much as possible beyond revisioning and other changes*
  - *Need to allow third parties to produce a translation of someone else's content*
  - *Ease of use to create content in any single language*
  - *Not multiple TLOs for multiple languages*
  - *Original text and its translations to be as close as possible*
  - *Content with mixed/multiple languages be received intact at the other end*

PROPOSALS

Text-based Proposal

- Baseline: Language code is NOT a property of objects, but a property of string itself

{ "type": "campaign", ...
"name": {"en": "Dridex Campaign - Botnet 121"},
"description": {"ja": "ボットネット 121 を活用する Dridex を元にしたキャンペーン"} ...}

MD5 Hash

{"type":"translation", ...
"text_ref: "41cb32a0d74d5d07f5362b3e66f245c9",
"ja ": "Dridex キャンペーン - ボットネット 121"},

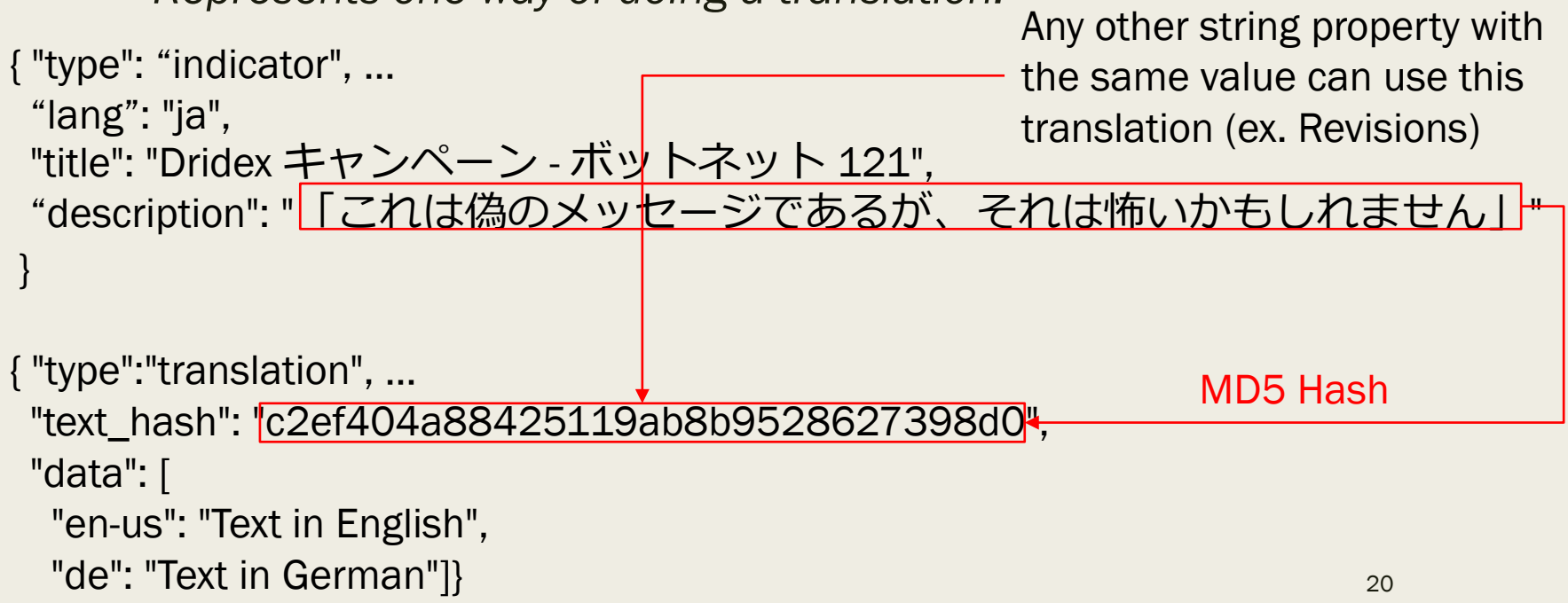
Any other string property with the same value can use this translation (ex. Revisions)

JK's take on it: Compact multiple translations

{ "type":"translation", ...
"fields":[
{ "text_ref: "41cb32a0d74d5d07f5362b3e66f245c9",
"ja": "Dridex キャンペーン - ボットネット 121"},
{ "text_ref": "e8465d411f6580e8b67d778f25a78234",
"ja": "ボットネット 121 を活用する Dridex を元にしたキャンペーン"},
{ "text_ref": "41cb32a0d74d5d07f5362b3e66f245c9",
"de": "Some German Title"},
{"text_ref": "e8465d411f6580e8b67d778f25a78234",
"de": "Some German Description"} ...]

Proposal - Bret

- No mixed language content in the main TLO,
  - *A UI or program could easily hide this from the user and the product would just send the content over the wire as an Indicator and a translation.*
  - *Allows third-party organizations to issue translations for any text found in any object and allows text to be translated once*
  - *Allows separate data marking to be applied to the translation object and external controls to applied.*
  - *Represents one-way of doing a translation.*



Proposal - JMG, modified from Bret

- Similar to Bret's proposal,
  - *Instead of text hashing, it is a similar to be TLO, providing the fields directly.*
  - *Provides the revision of the original object that was translated, so that when a new revision of the original object is received, you still may be able to use an old translation*

{ "type": "translation", ...
  "translation_of": ["indicator--a1201df6-c352-4a81-9c7c-5a6f896a4e31", 1],
  "created_at": "2016-05-31T05:43:23Z",
  "created_by_ref": "identify--9f127d43-505a-41d5-8743-ea3095c44fcf",
  "revision": 1,
  "lang": "en-us",
  "title": "Dridex 121",
  "description": "This is a description of the object in English."
}

Field-based Approach

- Simpler for single-language use cases and enables implementations that want something more complicated to do so.
- Approach:
  - *Top-level objects are single-language only. The language is specified by a lang tag that would be required on every TLO.*
  - *Translation objects translate entire TLOs, field-by-field*
  - *Allows for translation of non-string fields (open vocabularies, for example, which is an important use case)*


```
{ "type": "campaign",
  "id": "campaign--f44a8124-1ab1-4176-bcb8-bbc334d45228",
  "lang": "en",
  "title": "Shade of Palms",
  "description": "Shade of palms is a campaign that is bad", ...},

{ "type": "translation",
  "lang": "es",
  "translation_of": "campaign--f44a8124-1ab1-4176-bcb8-bbc334d45228",
  "title": "Sombra de las Palmas",
  "description": "Sombra de las palmeras es una campaña que es malo", ...}
```

Allan T's Suggestion

- Allow an object to be created natively in the language of the provider and connect the translated version of the object via relationships


```
{ "type": "campaign",
  "id": "campaign--a1201df6-c352-4a81-9c7c-
5a6f896a4e31", ,,,,
  "title": "Dridex Campaign - Botnet 121",
  "descriptions":
  "Dridex-based campaign leveraging Botnet 121",
     "intended_effects": [{"value": "theft-identity-theft"}
     ], "status": "Ongoing“ },
```

```
{ “type”: “relationship”,
   “relationship-name”: “translated-version”,
   “from”: “campaign--a1201df6-c352-4a81-9c7c-5a6f896a4e31”,
   “to”: “campaign--a1201df6-c352-4a81-9c7c-5a6f896a4e32”],
```

```
{ "type": "campaign",
  “lang”: “ja”,
  "id":   "campaign--a1201df6-c352-4a81-9c7c-5a6f896a4e32",
   "title": “Dridex キャンペーン - ボットネット ”,
  "description": " ボットネット 121 を活用する Dridex を元にしたキャンペーン ",
     "intended_effects": [{"value": " キャンペーン - "}
     ], "status": " 活用す “ },
```

Wouter's Options (1)

- Sub Option 1c is unique, putting "lang" information in the field name

Option 1

```
{ "type": "campaign",
  "id": "campaign--1234...1234",
  "lang": "en-us",
  ...
  "name": "bank attack 1",
  "description": "more information about bank attack",
  ...}
```

Sub Option 1a

```
{ "type": "translation",
  "id": "translation--9877...9877",
  ...
  "transaltion_of_ref": "campaign--1234...1234",
  "lang": "de",
  "name": "Bank Angriff 1",
  "description": "Weitere Informationen über Banküberfall"}
{ "type": "translation",
  "id": "translation--9866...9866",
  ...
   "transaltion_of_ref": "campaign--1234...1234",
   "lang": "fr",
   "name": "Attaque Bank 1",
   "description": "Plus d'informations sur la crise bancaire"}
```

Sub Option 1c

```
{ "type": "translation",
  "id": "translation--9877...9877",
  ...
  "transaltion_of_ref": "campaign--1234...1234",
  "name[de]": "Bank Angriff 1",
  "name[fr]": "Attaque Bank 1",
  "description[de]": "Weitere Informationen über Banküberfall"
  "description[fr]": "Plus d'informations sur la crise bancaire"}
```

Sub Option 1b

```
{ "type": "translation",
  "id": "translation--9877...9877",
  ...
   "transaltion_of_ref": "campaign--1234...1234",
   "name": {
     "de": "Bank Angriff 1",
     "fr": "Attaque Bank 1"}
  "description": {
    "de": "Weitere Informationen über Banküberfall",
    "fr": "Plus d'informations sur la crise bancaire"}}
```

Wouter's Options (2)

- Option 2 also puts "lang" information in the field name. This option allows you to serialize multiple translations of an object into a single Translation Object, as well as embedding translations in the main object.

Option 2

```
{ "type": "campaign",
  "id": "campaign--1234...1234",
  "lang": "en",  // default language used for all fields without explicit lang tag
  ...
  "name": "Bank Attack 1",
  "description": "more information about bank attack",
  "description[nl]": "meer informatie over de bankenaanval",
  ...}

{ "type": "translation",
  "id": "translation--9877...9877",
  ...
  "translation_of_ref": "campaign--1234...1234",
  "name[de]": "Bank Angriff 1",
  "description[de]": "Weitere Informationen über Banküberfall",
  "name[fr]": "Attaque Bank 1",
  "description[fr]": "Plus d'informations sur la crise bancaire"}
```

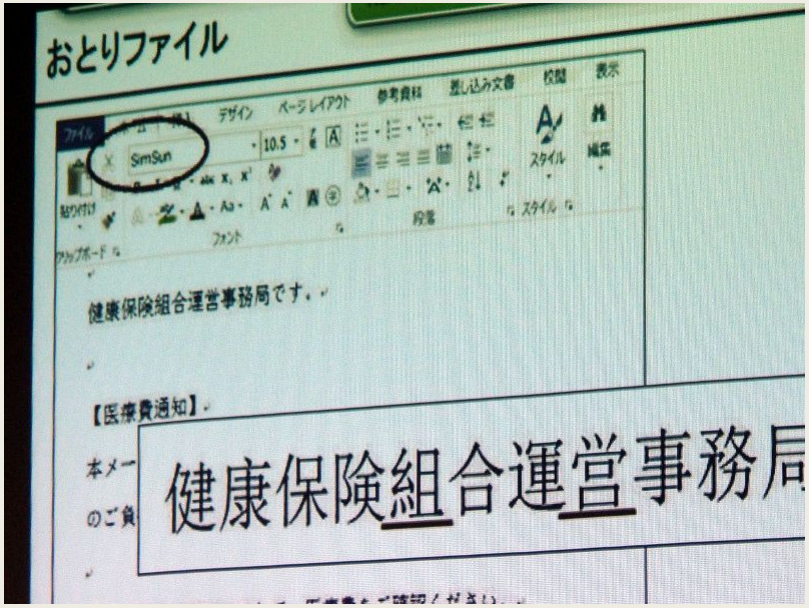
MISCELLANEOUS

References

- List of ISO 639-1 codes
  https://en.wikipedia.org/wiki/List_of_ISO_639-1_codes
- CJK characters
  https://en.wikipedia.org/wiki/CJK_characters
- The JavaScript Object Notation (JSON) Data Interchange Format (RFC7159)
  https://tools.ietf.org/html/rfc7159
- XLIFF Version 1.2
  http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html
- XORCISM - eXpandable Open Research on Cyber Information Security Management
  https://github.com/athiasjerome/XORCISM

Not Just Code, But Also Fonts …

- 「Emdivi」使い多発する日本への標的型攻撃、ラックが詳細報告 – ASCII - 20150617
- http://ascii.jp/elem/000/001/019/1019316/



おとりドキュメントのサンプル。一部の文字に日本語フォントではなく、Windows標準の中国語フォントである「SimSun」が使われている