Bridging Together Independent Islands

STIX Custom Objects and Matching Mechanisms to Correlate Cyberspace and Real-World Data

Toshitaka Satomi (not "Satomi Toshitaka")
Ryusuke (not "Ryuusuke") Masuoka
- Fujitsu System Integration Laboratories

Outline - Our Next Journey

1. Motivation
2. Initial Thought
3. Realization
4. Solution
5. Demo
6. Summary and Takeaway Messages

1. Motivation

- After FIRST2020, we continued our journey to widen CTI applications
- Toshi got a request from law enforcement (LE) practitioners
  - They wanted to correlate IP addresses and bank account information
- While thinking about the case, I realized this is common to other communities

2. Initial Thought

- Plain STIX has only cyberspace data objects, not real-world data
  - Need to use STIX custom objects
- MISP Standard has the “bank-account” and “person” objects
  - Can be exported in the STIX 2.1 format, using STIX Custom objects
- Toshi thought he can get away with it ...

- **... but it was not that simple**

3.1. Realization: *Who Am I!?*

- Ryu opened a bank account in Japan …
- Data to open an account in the US/Japan
  - They are conceptualized very differently like ***two independent islands***
    - MISP “person” object will do for the US
      – First name, middle name, last name
  - Matching first names would not be easy either

Name in Kanji

Last name

First name

Name in Katakana

Last name

First name

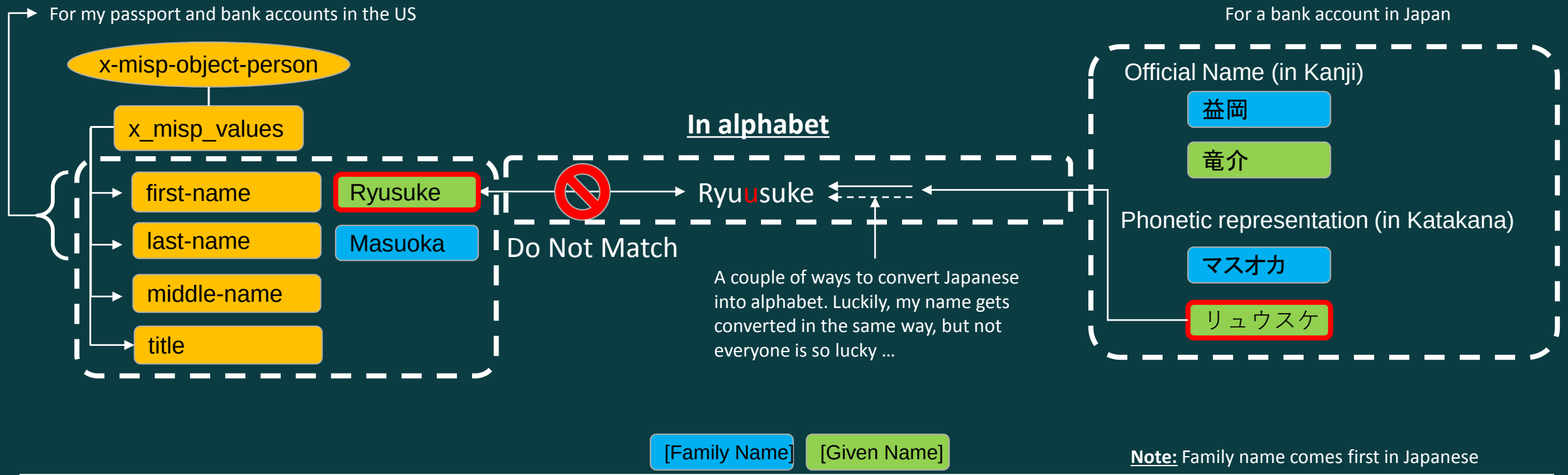
Bank account application form

[Tip] A Little Lesson on Japanese

- A variety of character sets used in everyday Japanese writing
  - **Kanji (Chinese Characters):**
    愛 意 上 得 終 漢 霧 空 …
  - **Katanaka (Phonetic – Double-byte/Zenkaku):**
    ア イ ウ エ オ カ キ ク …
  - **Katakana (Phonetic – Single-byte/Hankaku):**
    ｱ ｲ ｳ ｴ ｵ ｶ ｷ ｸ …
  - **Hiragana (Another Phonetic):**
    あ い う え お か き く …
  - **Alphabet (Double-byte/Zenkaku):**
    Ａ Ｂ Ｃ Ｄ Ｅ Ｆ Ｇ Ｈ …
  - **Alphabet (Single-byte/Hanakaku):**
    A B C D E F G H …
- My name in …
  - **Kanji (Chinese Characters):**
    益岡 竜介
  - **Katanaka (Phonetic – Double-byte/Zenkaku):**
    マスオカ リュウスケ
  - **Katakana (Phonetic – Single-byte/Hankaku):**
    ﾏｽｵｶ ﾘｭｳｽｹ
  - **Hiragana (Another Phonetic):**
    ますおか りゅうすけ
  - **Alphabet (Double-byte/Zenkaku):**
    Ｍａｓｕｏｋａ Ｒｙｕｕｓｕｋｅ
  - **Alphabet (Single-byte/Hanakaku):**
    Masuoka Ryuusuke

3.1. Realization: *Who Am I!?*

- Problems:
  1. Names are structured differently in the US and Japan
  2. Therefore, one needs to specify which field to match with which
  3. Matching (ex. first name) values is not simple either



3.1. Realization: *We have opened Pandora's box!*

3.2. Discussion

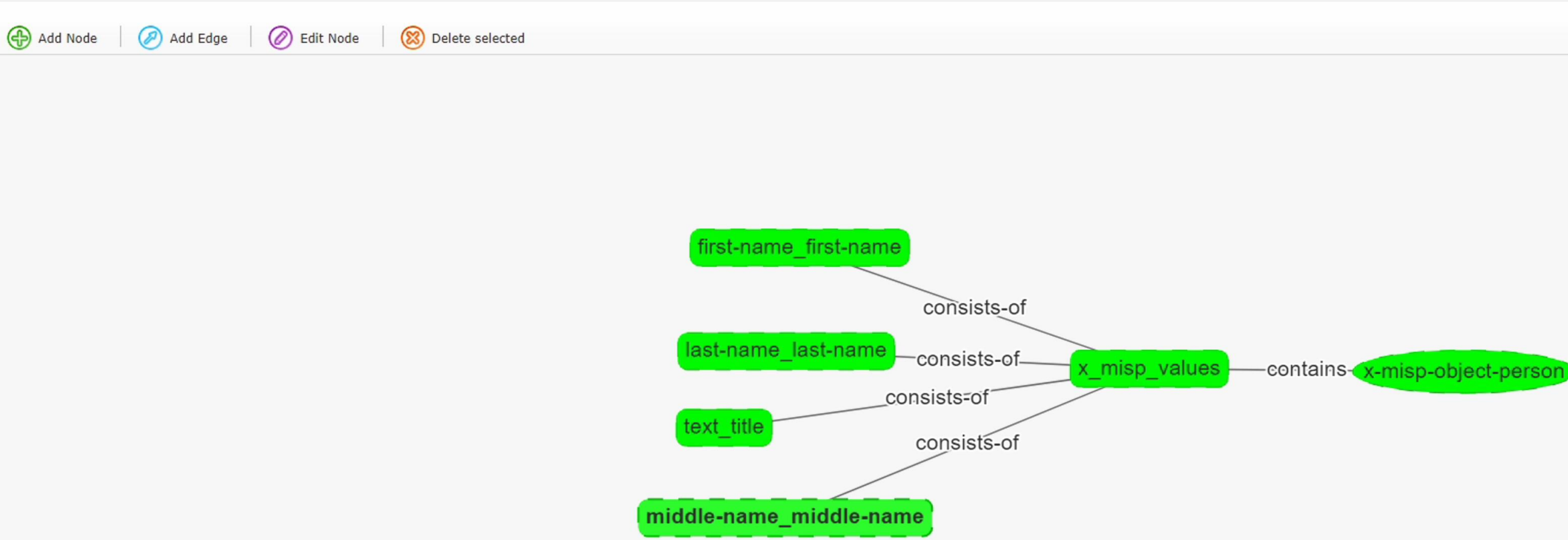
- Found many cases beyond Japanese names
  - Russian family names for male and female members (Livinsky/Livinskaya)
  - "cafe" and "café"
  - Company names in Japanese - Some have more than 40 variations
  - Aliases for threat actors and pieces of malware

- Need easy-to-create custom objects and flexible matching mechanisms to deal with real-world data

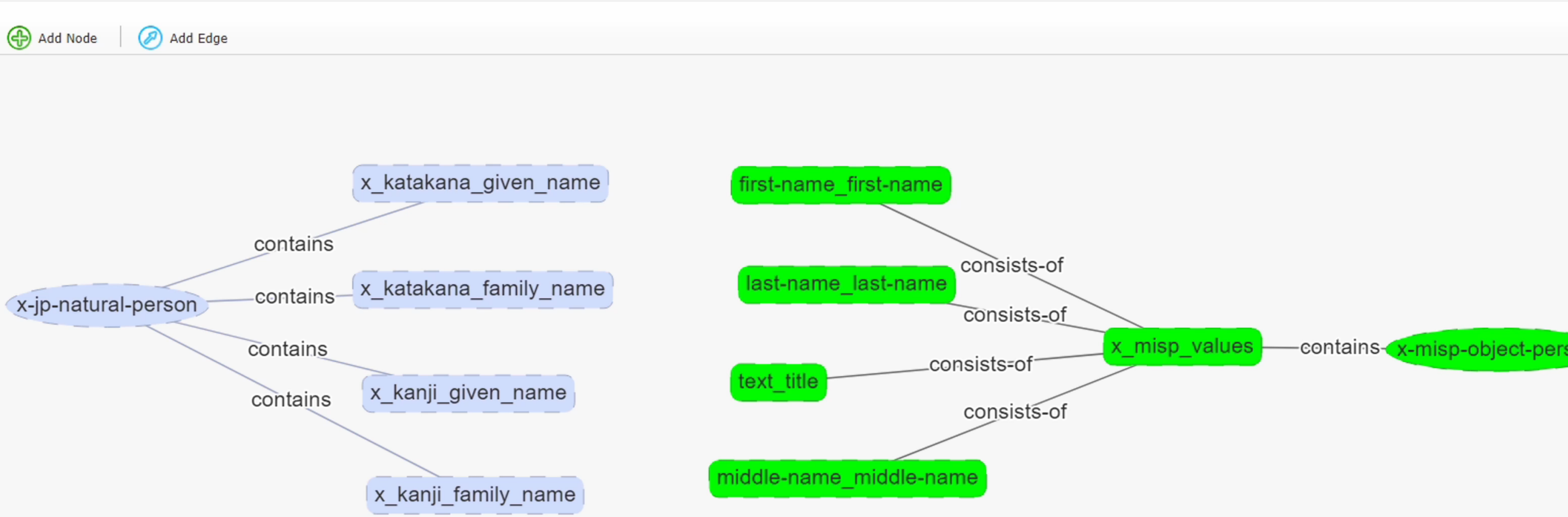
4. Solution – STIX Customizer and Matching Mechanisms

- ... implemented in our TIP, S-TIP
  - STIX Customizer
    - Easy to create STIX Custom Object Models
  - Explicit Correlation
    - Limiting which properties are matched
  - Fuzzy Matching
    - Absorbing notation fluctuations
- Demonstration of how those S-TIP mechanisms are utilized to make meaningful matching of real-world data

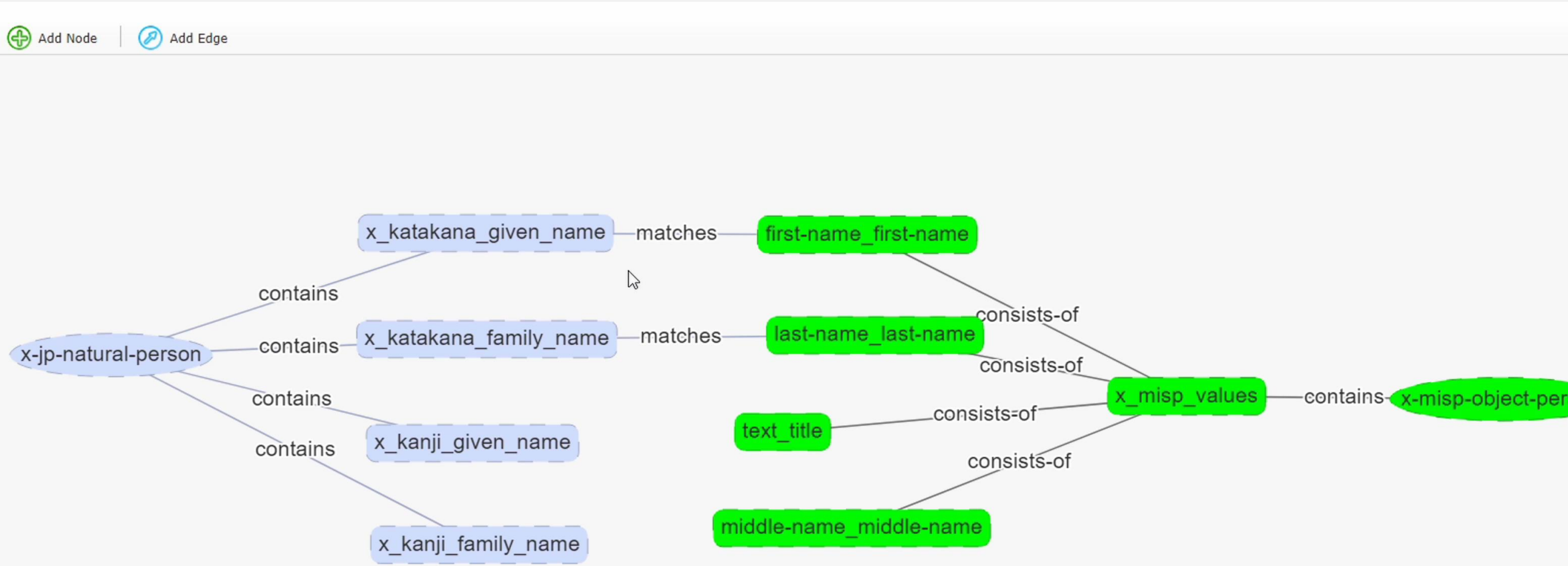
5. Demo (Create Custom Object and Properties)



5. Demo (Define Explicit Correlation)



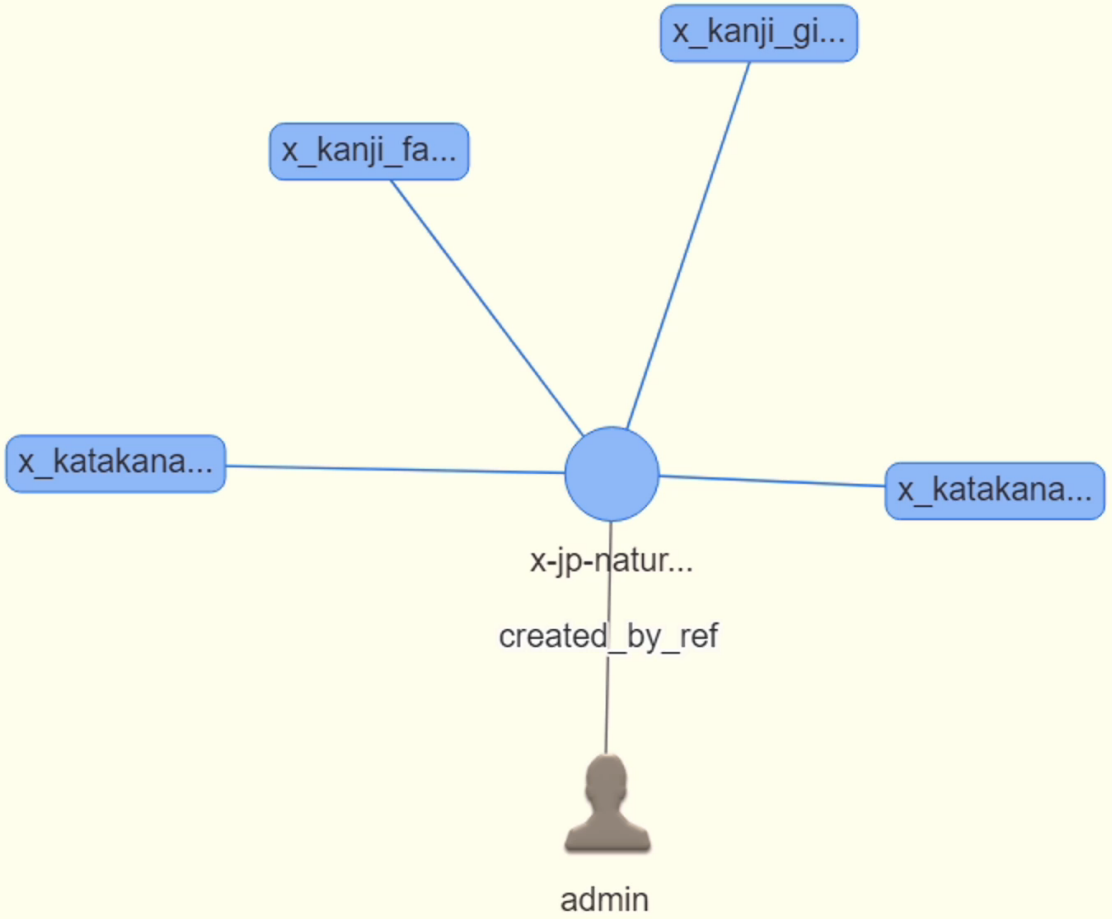
5. Demo (Define Fuzzy Matching)



5. Demo (Explain the demo Data)


```
    "spec_version": "2.1",
    "x_misp_values": {
      "text_title": "Dr.",
      "last-name_last-name": "Masuoka",
      "first-name_first-name": "Ryusuke"
    },
    "created": "2022-06-01T01:28:20Z",
    "modified": "2022-06-01T01:28:20Z",
    "labels": [
      "misp:type=¥"person¥"",
      "misp:category=¥"misc¥"",
      "misp:to_ids=¥"False¥"",
      "from_object"
    ],
    "x_misp_category": "misc",
    "created_by_ref": "identity--42614707-a1f5-4ae9-94e5-3d7cd3c7c6d9",
    "type": "x-misp-object-person"
  }
]
```

5. Demo (Graph Analytics)

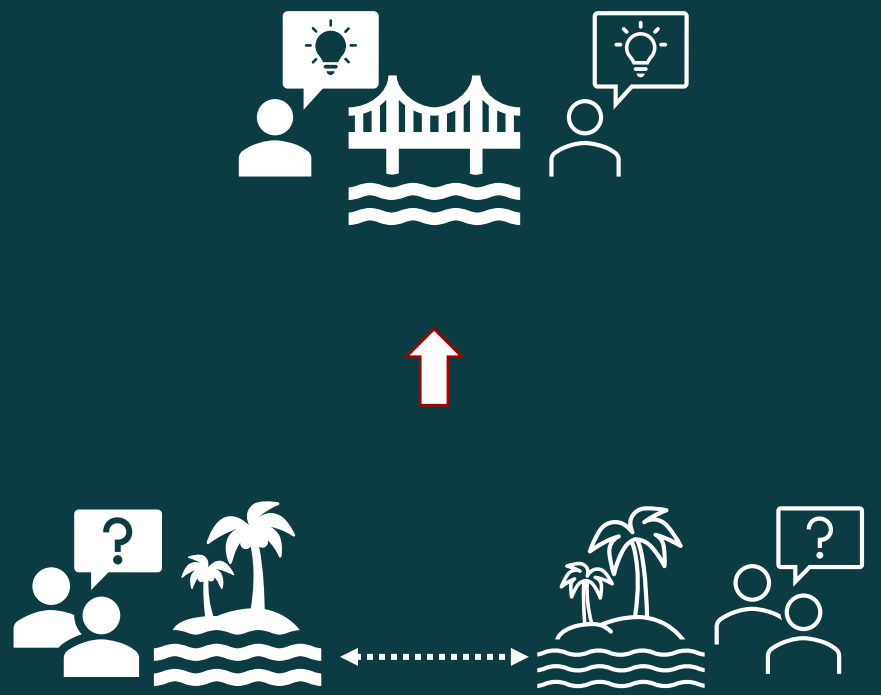


6. Summary and Takeaway Message

- STIX Customizer and matching mechanisms proposed and implemented in S-TIP
  - "STIX Customizer" to easily create STIX custom objects
  - "Explicit Correlation" and "Fuzzy Matching" to make meaningful matching possible for real-world data
- Future work
  - Apply these technologies to problems from the LE practitioners … and others!
  - Faster "fuzzy matching"
  - Study and deal with other notation fluctuation cases
  - Enable sharing STIX Custom Object schemas and matching rules

6. Summary and Takeaway Message

- Easy-to-create custom objects and flexible matching mechanisms to utilize real-world data like independent islands along with cyberspace data



Thank You!!!

- Toshitaka Satomi

 @stmtstk

- Ryusuke Masuoka

 @rmasuoka

We want contributors!!
https://github.com/s-tip/

S-TIP