How to Create Effective Structured Intelligence Extensions for TIPs

Peter Ferguson
CTI Specialist, EclecticIQ
Who am I

• Cyber Threat Intelligence Specialist, EclecticIQ
• Previously responsible for designing extensions for the EclecticIQ Intelligence Centre
• Now doing threat research for EclecticIQ's Intelligence & Research group
Aim of the Talk

• Convey lessons learned from designing extensions for a threat intelligence platform
• Define requirements needed to start building structured intelligence extensions
• Detail the key goals of an extension
• Help analysts / engineers create extensions that work for them
Key Concepts
What is a TIP?

A centralised platform to ingest, normalise, correlate and analyze threat data from various sources to help support defensive operations.
What is Structured Threat Data?

Structured Threat Data is information about cyber security threats that is consistent and machine readable. This is to improve system-to-system sharing, threat analysis and automation.
What are Extensions?

An extension is an add on to the TIP that allows for the ingestion of data from an external source or platform into the TIP; e.g. an incoming feed or an enricher.

Extensions exist to transform incoming data into a single consistent model and to provide flexibility for the TIP.
Define Your Data Model

- How do you want to represent threat data?
- What are the limitations of the platform you use?
- Can get inspirations from open source standards (STIX)
The Data Life Cycle

**Ingestion**
- Is data ingested within a reasonable timeframe?
- How is data deduplication being handled?

**In-Platform**
- Is the data normalised?
- Is the data searchable?
- Is the data walkable?
- Is the data classified correctly (confidence etc.)?

**Dissemination**
- Is data classified correctly?
Ingestion
Source Data Including IDs

```json
{
  "malware_data": [{
    "sample_1": {
      "id": "1234abcd",
      "type": "lok",
      "ip": "63.250.39.66",
      "malware": "emotet",
      "tags": ["malware", "windows", "malicious", "downloader", "c2"],
      "timestamp": 1663692836,
      "source": "https://www.eclecticiq.com/
    },
    "id": "5678ertq",
    "type": "lok",
    "hash": "9e63215451ebab2a13e16404a27ab97bb507f3668473826f5964fa2238158c38",
    "malware": "emotet",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com/
  },
  {
    "id": "1234abcd",
    "type": "lok",
    "ip": "63.250.39.66",
    "malware": "emotet",
    "tags": ["malware", "windows", "malicious", "downloader", "c2"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com/
  }
}
```
Source Data Without IDs

```json
{
  "malware_data": [{
    "sample_1": [{
      "type": "loc",
      "ip": "63.250.39.66",
      "malware": "emotet",
      "tags": ["malware", "windows", "malicious", "downloader", "c2"],
      "timestamp": 1663692836,
      "source": "https://www.eclecticiq.com/"
    }]
  }
}
```
How to Fix - Source Data Without IDs

```json
{
  "malware_data": [{
    "sample_id": {
      "type": "locl",
      "ip": "63.250.39.66",
      "malware": "emotet",
      "tags": ["malware", "windows", "malicious", "downloader", "c2"],
      "timestamp": "1663692836",
      "source": "https://www.eclecticiq.com/"
    },
    {"type": "locl",
      "hash": "8e63215451e8ab2a13e16494a2fcb97bb5075f36684738260f564fa223816c38",
      "malware": "emotet",
      "tags": ["malware", "windows", "malicious", "downloader"],
      "timestamp": "1663692836",
      "source": "https://www.eclecticiq.com/"
    },
    {"type": "locl",
      "ip": "63.250.39.66",
      "malware": "emotet",
      "tags": ["malware", "windows", "malicious", "downloader", "c2"],
      "timestamp": "1663692836",
      "source": "https://www.eclecticiq.com/"
    }
  }
}
```
Targeted Collection

```json
{
  "search_result_data": [{
    "type": "doc",
    "hash": "b2f58eda1180aa69c80d68d078b29f62c21ebd9cdc18266a46fbd991e3f36a88",
    "malware": "emotet_document",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com",
    "search": "behaviour_network:https://213.239.212.5:443"
  }]
}

{
  "search_result_data": [{
    "type": "doc",
    "hash": "b2f58eda1180aa69c80d68d078b29f62c21ebd9cdc18266a46fbd991e3f36a88",
    "malware": "emotet_document",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com",
    "search": "content:6F 6E 22 2C 22 52 4C 44 46 77 6E 6C 6F 61 64 54 46 69 6C"
  }]
}
```
In-Platform
Making Assumptions on Data

```
{
  "data": [{
    "type": "ioc",
    "hash": "994c66e6d07592ce6a62bd2b667c60694e862f17f7e74088feb8d964595f18ba",
    "threat": "emotet",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com/",
  },
  {
    "type": "ioc",
    "hash": "bed236e8cf7cbb0a79f4620f4fb84796269a3eb8ec332b7b64375e1ed9627a",
    "threat": "phishing",
    "tags": ["phishing", "malicious"],
    "timestamp": 1663701184,
    "source": "https://www.eclecticiq.com/",
  }
}
```
How to Fix - Making Assumptions on Data

1. Create Logic
2. Flatten
Leverage Platform Automation Capabilities

```json
{
  "data": [{
    "type": "ioc",
    "hash": "994c6b666d07592ce5a62d2b667c60694e862f17f7e74088feb8d964595f10ba",
    "malware": "emotet",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "industries": ["financial", "ecommerce", "healthcare", "Academia"],
    "actor_type": ["criminal"],
    "timestamp": 1683692836,
    "source": "https://www.eclecticiq.com/"
  }
}]
```
Too Much Related Data
How to Fix - Too Much Related Data
### Consistent Classification

A consistent classification system is crucial for ensuring that all classifications are applied uniformly. This helps in avoiding confusion and ensures that all stakeholders understand the criteria for classification.

#### Classification vs. Number

<table>
<thead>
<tr>
<th>Classification</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown</td>
<td>N/A</td>
</tr>
<tr>
<td>None</td>
<td>0</td>
</tr>
<tr>
<td>Low</td>
<td>1-33</td>
</tr>
<tr>
<td>Medium</td>
<td>34-66</td>
</tr>
<tr>
<td>High</td>
<td>67-100</td>
</tr>
</tbody>
</table>

The table above provides a clear mapping of different classifications to their respective numeric values. This helps in quickly assessing the severity level of a particular classification.

Example data:

```
{
  "data": [{
    "type": "ioc",
    "sha256_hash": "994c6b6e6d07592ce4a62bd2b667c6b694e862f17f7e74088fbeb9d964595f10ba",
    "sha1_hash": "21288cb8d696d79f8e9bb9966f17f7e74088febeb9d96595f10ba",
    "md5_hash": "51b3e08cb5b18f4d8676b4a9bebb0fd8",
    "confidence": "50%",
    "malware": "emotet",
    "tags": ["malware", "windows", "malicious", "downloader"],
    "industries": ["financial", "ecommerce", "healthcare", "Academia"],
    "actor_type": ["criminal"],
    "timestamp": 1663692836,
    "source": "https://www.eclecticiq.com/
  }
}
```
Dissemination
Extension Tips for Dissemination

• The bulk of dissemination is automated
• Exact process will be platform dependent
• All platforms will use some kind of filtering
• This stage is reliant on good design decisions in the earlier stages
Enrichers
Granular Enrichers

What is the infrastructure used to download this file?
Enricher in Use
Key Goals & Tips
Goals & Tips

Key Goals
- Ingestion of data within a reasonable timeframe
- Normalisation of data into a single model
- Correct classification of data
- Enrichment is granular

Tips to takeaway
- Keep it simple
- If the platform can do it, let it
- Avoid assumptions on data
- The extension is only as strong as its weakest link
- Extensions are about making the correct trade-offs between all stakeholders
Any Questions?

blog.eclecticiq.com