TLP:WHITE

INTEGRATION OF INFORMATION IN OPENCTI

DOCTRINE OF USE OF ANSSI'S CTI TEAM

1.0

07/01/2021

TLP:WHITE

Table of contents

1 Using OpenCTI at ANSSI: goals and requirements

1.1 Context and objectives

The following document presents why and how the Cyber Threat Intelligence (CTI) team at the French National Agency for the Security of Information Systems (ANSSI) integrates data in OpenCTI today.

This doctrine has been published in order to help entities and people interested in the plateform understand how OpenCTI can be leveraged to integrate and enrich information and to investigate and export knowledge. The following models and use cases are the ones implemented by the CTI team at ANSSI. **They do not represent a general guideline or standard for using OpenCTI or for storing and enriching CTI knowledge in general, but one of the possibilities**.

At ANSSI, the daily mission of the CTI team is to collect, integrate, process and analyze information on cyberattacks and intrusion sets, which are or may be targeting France and French interests. One of the goals of the team is to anticipate and help mitigate cyberattacks against French interests. It is done by sharing consolidated knowledge with other teams within ANSSI, with ANSSI's constituants, partners and the broader community.

The initial requirements for developping OpenCTI and the reasons the CTI team at ANSSI uses the platform today are the following:

- integrate and store in a structured format information about cyberattacks and intrusion sets (especially TTPs and victimology);
- enrich, vizualise and investigate this data to create knowledge;
- extract and share this knowledge in various formats, while adequately dealing with different levels of marking.

It is important to note that the chosen approach of ANSSI's CTI team in using OpenCTI differs from the approaches adopted by others in the community. The main reasons for making these choices at ANSSI were the available resources, the use of other tools (such as MISP for storing signatures) and specific goals in using the platform.

This document was first written to help the analyst and ensure consistency in the way data is being structured, as the OpenCTI database used at ANSSI is filled manually by analysts[1]. This means that it is up to the analyst to decide which document he/she wishes to import, and to manually extract and integrate the information. Therefore, the choice was made to limit both the quantity of reports to process and the type of data to extract. The criteria of selection are listed in the following sections.

This doctrine will be modified as the objectives at ANSSI in using the platform evolve, and with the technical changes which are continuously made in the platform.

This document does not yet integrate all the features brought by OpenCTI v4. Some features still need to be tested with concrete usecases in order to establish how they could best inserted in this general doctrine. The document will be updated accordingly.

1.2 General criteria for choosing a document to import

A document[2] has to fullfill the following checklist in order to be imported in the platform by the analyst:

- it contains information on one of the types listed below: an intrusion set, a campaign, an incident, TTPs, victims (organizations, sectors, countries), a malware or other technical information;
- this information is related to an intrusion set, or to a set of adverse activities already followed by ANSSI, or targeting French interests, or a sector of interest to ANSSI;

[1] However, one current project is automating integration.
[2] Here defined as any type of document and format which can be imported in OpenCTI.

- it contains more precise or new knowledge compared to what has already been processed[3]. In most cases, a relation will therefore will only be used once. If an information has its marking changed, it is processed as a different information. For instance, the use of a TTP by an intrusion set documented first in a report with a "TLP:AMBER" marking, then made public in a blog post, is considered as a new information as it requires creating a new relation with the corresponding marking level.

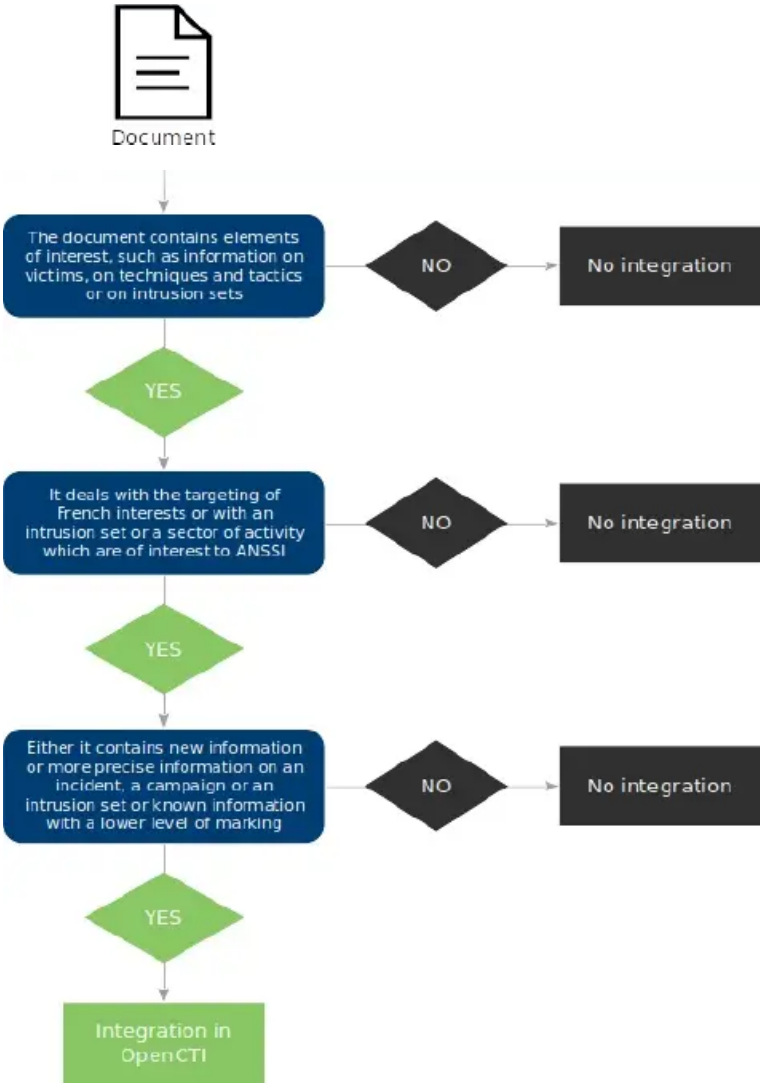
When integrating information in OpenCTI, it is important from the start to have a clear picture in mind of what exploitation and what outputs are wished for, in order to extract the correct data and use the right model for structuring the information.

In the context of the specific use of OpenCTI made by ANSSI's CTI team, especially because integration is done manually, it is advised to read carefully the documents in order to identify the data which has to be extracted and how to structure it **before** actually starting integrating it in the platform. This approach can be immensely time-saving for the analyst.

When possible, it is also recommanded to favor precise documents on one-time incidents and attacks over more general reports, for instance on activities of an intrusion set over several years, with which it is usually more difficult to identify precise incidents and timelines.

The following graph illustrates the decision path described above.

[3]The goal in using this approach is to gain time and avoid integrating twice the same information. In the case of this doctrine, the difference of source is not considered a new information. However, this can be adapted, for instance to follow when and what a security editor has published about an intrusion set or a campaign



1.3 Required information types for the models

As of today, it was decided to limit integration to information belonging to the types listed below. **Observables and indicators are not imported in the CTI team's OpenCTI instance**. However, for clarity, indicators were included in one model (see section 2.2.3).

This section only deals with which information types are collected for integration in OpenCTI. How to modelize the information and organize the entities is described in section 2.2.

Datasets are currently being implemented for some of these types in the instance. The list can be found in appendix 4.1.

It is strongly advised to use the description fields for all information which cannot be structured or will not be used in a stuctured way for outputs.

Victimology

Information on victimology should be as precise as possible: the exact victim (organization or person), or at least the sector of activity and the region or country to which it belongs, if information about the victim are too imprecise.

If a victim is anonymized and only loosely identified (for instance "*A French company in the retail sector*"), the incident type should be used instead of creating an unamed entity with the organization type. Integrating precise information on the attack against this entity (dates of attack, activity sector and location, tools, malware and TTPs, link to a campaign etc.) will be easier and clearer (see model in section 2.2.3).

TTPs

The MITRE ATT&CK and PRE ATT&CK matrices are implemented as reference bases for identifying and integrating TTPs in OpenCTI. The description field **has to** be used to store additionnal information, such as on the way the TTP was used in the context. For instance, for the "*Data encrypted*" TTP, it would be relevant to indicate the type of encryption which is used or characteristics in the implementation.

Adding text in the description field, especially for TTPs, is useful as it will give a better understanding and vision of how the intrusion set uses the TTP over time. This can be leveraged during incident response for example.

In addition, these descriptions will be visible in the kill chain view and in the different tables summarizing information, making the outputs more understandable (for more information on expected outputs, see section 3).

Malware and tools

In the platform, a distinction is made between malware and legitimate tools.

Disclaimer: the following definitions are only given to help the reader understand what is defined as a malware and as a tool in the context of this doctrine.

- **Malware** are defined as tools developped by or for threat actors and used by intrusion sets to achieve their objective on their target;
- **Legitimate tools** are defined as tools or services running on a machine or in a network and used to complete tasks necessary to maintain, run or secure the machine or the network. These tools can have their use or function hijacked by the intrusion set to fullfil its goals.

When malware or tools are mentionned in a report and do not already exist in the OpenCTI instance, their name and their main characteristics (such as language and main functions) should be registered if they are known.

If necessary, different versions of a malware or of a tool can be created by generating one entity for each version and linking them with a "variant of" relation. For instance, this could be interesting for malware such as BlackEnergy, which has several known versions.

Intrusion sets

If an intrusion set is identified as possibly responsible for, or linked to the malicious activity described in the report, the information is to be integrated.

Identifying intrusion sets is central in the platform and in the chosen data model. One of the goals in using OpenCTI is to structure and consolidate through time coherent knowledge on intrusion sets and their TTPs, victimology and malware.

Campaigns and incidents

Disclaimer: the following definitions are only given to help the reader understand what is defined as a campaign and as an incident in this context. It takes into account that the definition of a campaign will vary

with the intrusion set.

A campaign is defined as *"a set of malicious activities or attacks (sometimes also called a "wave of attacks") taking place within a time frame, against a set of victims, associated to a specific intrusion set and characterized by the use of (almost) identical versions of one or several malware"*.

The campaign can already have a public name given by a security editor, the community or the press, in which case it is strongly advised to reuse this name while creating the corresponding campaign entity, eventually adding a victim name and a date for clarity (for instance, "NotPetya - World - 2017").

The analyst can also assess, when investigating several attacks documented in different reports, that they are in fact part of the same campaign and create the entity to link these different activities. Criteria for defining what is a coherent campaign can change from one intrusion set to another and depends on the analysis.

The use of a campaign entity is interesting as it helps distinguish several waves of activity which are similar but distant in time. For instance, a report indicates that an intrusion set targeted a given sector of activity in a country A and the same sector but much later and in a country B. In that case, it can be useful to create two different campaigns, which will make the difference of temporality and location more visible.

An incident is defined as *a set of malicious activities associated with an intrusion set and targeting a precise victim for a continuous and identified time period*. As mentionned above in section 1.3, using the incident entity can be convenient to integrate information on an attack for which the victim is described but not clearly named.

2 Rules for proper integration

2.1 General rules

2.1.1 Working language

In order to facilitate exchanges with partners and to be able to automate STIX exports from OpenCTI, the choice was made to use **English for all the fields, including the description fields** .

2.1.2 Rules for the creation and management of entities

Although some datasets are implemented, it can be necessary to create new entities, for instance new organizations. Rules are necessary in order to ensure data consistency in the platform and for exports.

If in a dataset one entity is missing or has missing or wrong information, the error should preferably be reported directly in the GitHub repository of the project, in order to edit the original dataset.

For type of entities for which there is no preliminary dataset, if one new entity has to be created, the steps are the following[4]:

- whether the entity already exists or not should be checked (checks should be made especially for spelling discrepancies). If the entity already exists under a different name, adding an alias can be preferable;
- if it does not exist, the entity should be created with the proper type. **Description of the entity and marking are mandatory**. Aliases can be added to facilitate finding the entity later and to avoid creating duplicates (for instance, when creating the entity "ANSSI", adding the "CERT-FR" alias can be useful);
- when creating a campaign or an incident, the name which already exists in open source or in the report should be used. If there is none, the new name should be as precise as possible, because if it is too generic, such as "Campaign of Clop ransomware against Asia", there is a high risk of mixing up different campaigns. The

[4]These steps have been written down for a case in which an analyst is manually creating the entity. However, these rules can help set up an automated process for creating new entities

suggested naming convention (depending on the available information) is "targeted entity, sector or country - intrusion set identified - month and year of targeting". This would give for instance "Governmental entities in South Korea - Clop ransomware - 2019".

To have a clearer picture of which entities already exist and avoid duplicates, navigating in the different tabs in the platform can be useful. OpenCTI also integrates an automatic check for duplicates when creating a new entity, and there is a dedicated tab for merging duplicates.

To find entities and especially TTPs using the research field, it is necessary to use double quotation marks. Refering to the official MITRE website can help a lot in finding the right TTP.

2.1.3 Understanding, creating and managing relations

Reminder on how relations work

As it is detailed in the general documentation on the platform, which you can find at `https://www.notion.so/OpenCTI-Public-Knowledge-Base-d411e5e477734c59887dad3649f20518` , relations in OpenCTI have three interesting and very useful characteristics:

- the type of relation between two entities depends on two factors: the type of both entities and the direction in which is drawn the link. The full list of relation types is available in the official documentation of the project;
- inference rules can be set. For instance, if a campaign has several incidents and *is attributed to*[5] an intrusion set, then the incidents will also be "attributed to" the intrusion set.

Rules for creating relations

When creating new relations between entities, some rules also apply, which are specific to this use case:

- whether the relation does not already exists should be checked (if it does, it will be suggested);
- in the case in which there is a already existing relation matching the data extracted from the report:
  - if there is no new information in the document which could enrich the exiting relation, data should not be integrated (more information in the graph section 1.2);
  - if there is new information which can enrich the relation (new or more precise dates, enhanced description etc.), they are to be added. The different sources of the relation will be visible. However, if there is a change in the marking level, a new link has to be created to avoid future leaks when exporting data.
- if a new relation has to be created, several fields will be filled by default based on the information from the source document. Some information can be added or edited:
  - **the date:** the "last seen" and the "first seen" are automatically filled with the source document's dates. However, if more precise activity dates are available in the document, the field should be modified accordingly. Dates can only be on a YYYY-MM-DD format. Therefore, if the document only gives an imprecise date (for instance only a month or a part of the year), the 1st of the month should be put for the day and the tag for "imprecise date" should be added (more information in section 2.1.4);
  - **confidence level:** there are today four confidence levels: low, moderate, good and strong (for details on these levels, refer to appendix 4.2. The confidence level of the relation is by default the confidence level of the source document. If the source document has no confidence level, it is automatically set on "low";
  - **description:** the description field is free and unstructured. It has to be filled in English and the analyst must use it to add important unstructured information which cannot be put in other fields or translated into an entity and a relation.

[5] Here, "attributed to" is the relation type in the OpenCTI platform. Using this relation type does not mean that a formal, political attribution is made.

2.1.4 Use of tags

Tags can be added depending on needs or if the content of the document fits identified use cases. The following tags are implemented in the OpenCTI instance used by ANSSI's CTI team.

Information type	Tag value
Type of actor	Cybercriminal
Type of actor	Hacktivist
Type of actor	State-sponsored
Type of actor	Mercenary
Type of actor	Patriotic hacker
Type of actor	Employee
Type of actor	Competitor
Motivation	Espionnage
Motivation	Sabotage
Motivation	Revenge
Motivation	Financial gain
Motivation	Disruption
Motivation	Neutralization
Motivation	Divulgation
Status of attack	Targeted
Status of attack	Compromised
Status of code	Shared code
Status of code	Unique code
Status of the entity	Victim
Status of the entity	Attacker
Status of the date	Precise date
Status of the date	Imprecise date
Status of the date	Unknown date
Type of code	Loader

Type of code	RAT
Type of code	Stealer
Type of code	Ransomware
Type of code	Wiper
Type of code	Cryptominer
Type of code	Exploit

2.2 Modelling information in OpenCTI

2.2.1 Preliminary remarks

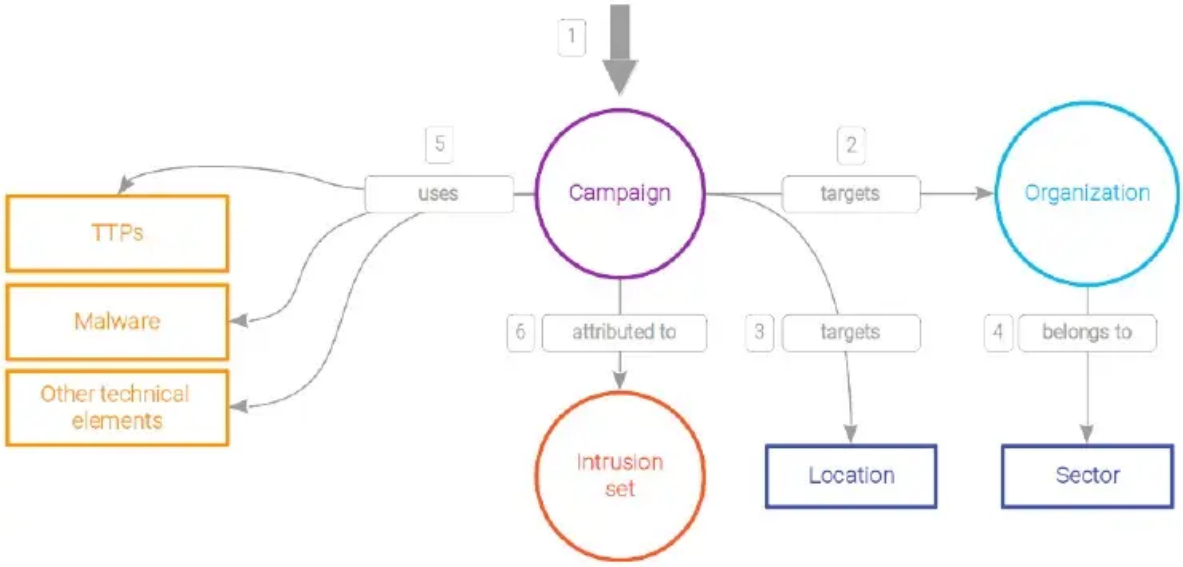
The following graphs aim at illustrating ANSSI's CTI team's datamodels for structuring information in OpenCTI. These models help uniformize the integration process, define inference rules and how data is visualized and exported. They indicate the direction in which relations have to be drawn and the expected relation type. As said above, the team has chosen not to integrate observables and indicators. However, one alternative model with the integration of indicators is presented.

In some cases, these models can be combined: for instance, a report can deal with an intrusion set and give both precise information about attributed campaigns and incidents (in which case both models in 2.2.2 and in 2.2.3 are to be used) and can also list TTPs and malware more generally used by the intrusion set (in which case the model in 2.2.4 is also to be used).

The same model can be replicated several times within the same report: for instance, several incidents and their details can be attributed to one campaign and several campaigns can be attributed to one intrusion set. A report can also deals with several intrusion sets at the same time. One victim organization can be targeted by several incidents at the same time: the identified use case could be of a response incident operation during which several intrusion sets are identified. In that case and if possible, an incident should be created for each intrusion set which has been identified. A generic incident can be created for TTPs and malware which has been identified but not attributed to a specific intrusion set yet.

If information about the victim are not precise enough to create the matching entity, the activity sector and the location should be directly linked to the central entity (the incident, the campaign or the intrusion set entity), if they are known.

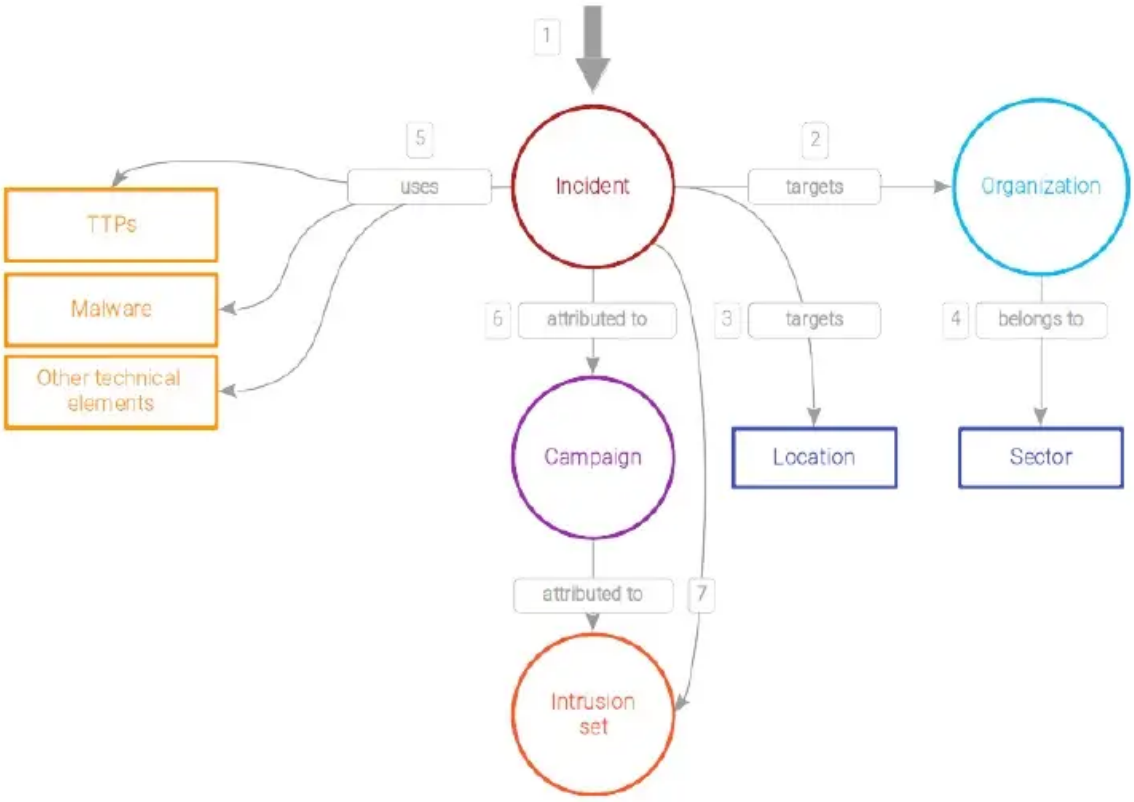
2.2.2 Model for a campaign



1. The selected document deals with an attack campaign...
2. ... for which one or several victims have been identified.
3. The campaign targets one region or one country...
4. ...and the victim belongs to a specific sector of activity.
5. Technical elements have been identified and can be linked to the campaign.
6. The campaign can be linked to an intrusion set.

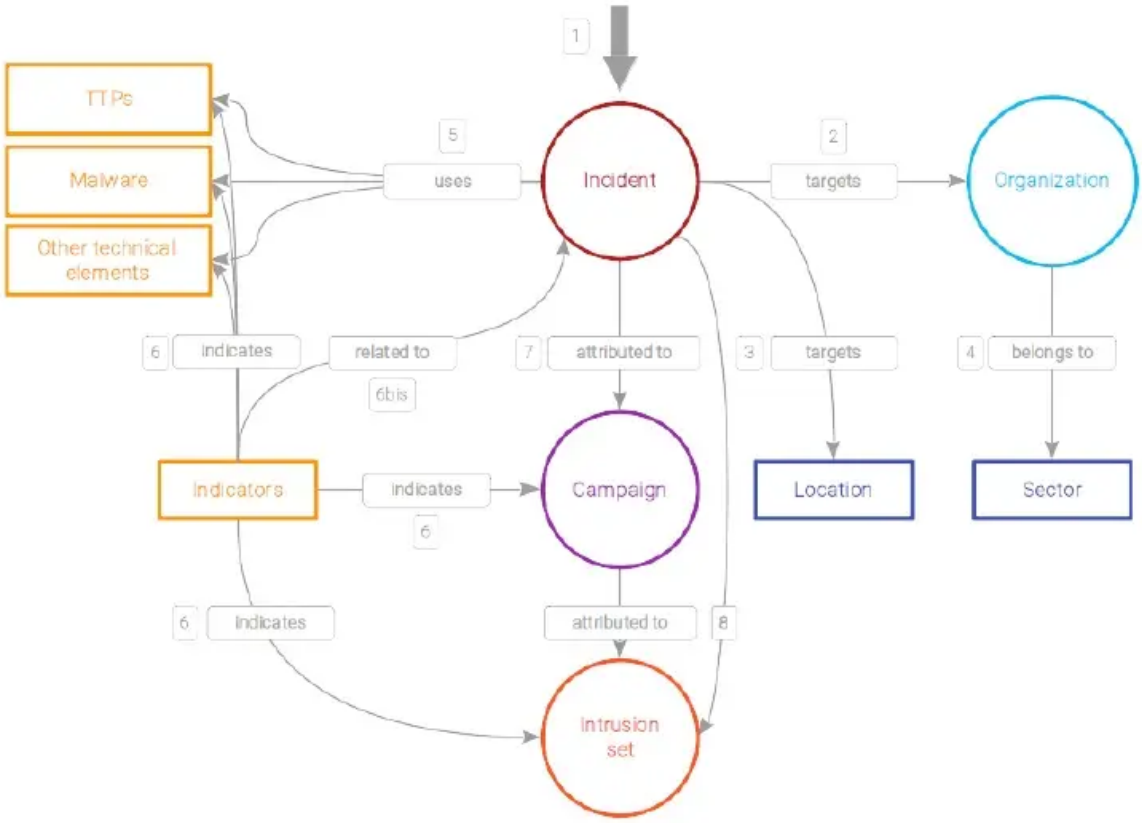
2.2.3 Model for an incident

Main model



1. The selected document deals with an incident...
2. ... for which one or several victims have been identified.
3. The incident targets one region or one country...
4. ... and the victim belongs to a specific sector of activity.
5. Technical elements have been identified and can be linked to the incident.
6. The incident may belong to a larger campaign.
7. The campaign is linked to an intrusion set. The incident can also be directly linked to the intrusion set.

Alternative including indicators



1 The selected document deals with an incident...

2 ... for which one or several victims have been identified.

3 The incident targets one region or one country...

4 ... and belong to a specific sector of activity.

5 Technical elements have been identified and can be linked to the campaign.

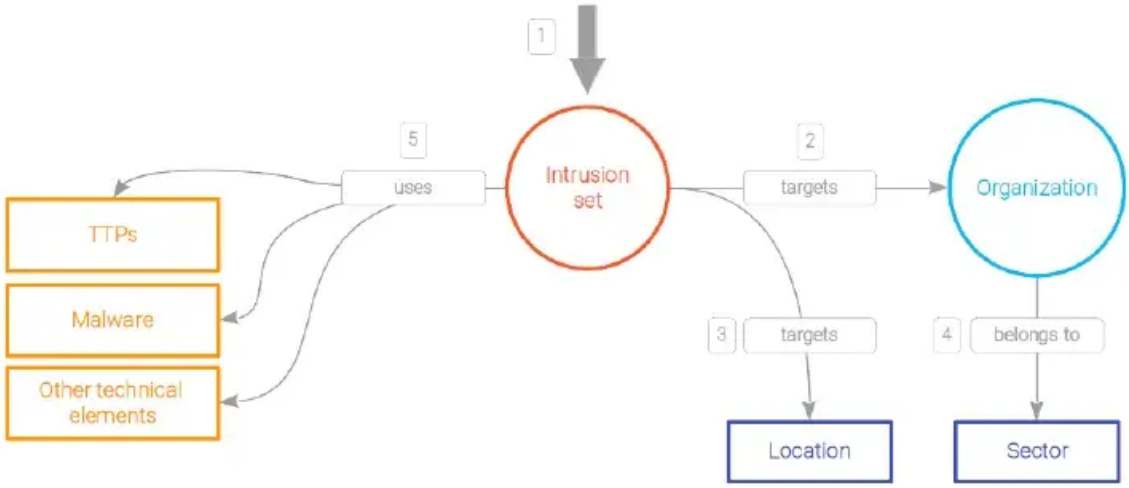
6 Indicators may be linked to the relation between the incident and the technical elements, or directly to the campaign or the intrusion set.

6bis They also may be related to an incident directly.

7 The incident may belong to a larger campaign.

8 The campaign is linked to an intrusion set. The incident can also be directly linked to the intrusion set.

2.2.4 Model for general information on an intrusion set



1. The selected document deals with general knowledge on an intrusion set...
2. ... which targeted one or several identified victims...
3. ...and one region or one country...
4. ... and belong to a specific sector of activity.
5. Technical elements have been identified and can be linked to the intrusion set.

3 Expected use cases

Defining what data is integrated and how it should be structured in OpenCTI is helpful for obtaining outputs matching expected use cases. The format of the output (meaning what the plateform allows for visualization and exports) should be distinguished from the desired use case (meaning what the team would like to visualize and export once data has been integrated).

Today, the identified and desired use cases for ANSSI's CTI team are the following:

- analysts in the CTI team or from other teams at ANSSI are able to vizualise and understand information directly in the platform on a subject they have not been following before, or on a subject they followed some time ago;
- information from the "activity" tab help other analysts and managers to follow and export data on the analysts' activity;
- analysts in the CTI team are able to investigate this knowledge and relations between things;
- information (either the raw information structured as a table, or as an image) on kill chains or lists of TTPs, malware, victims etc. can be extracted to be embedded in a written report;
- exports can be generated in JSON or CSV formats and shared with partners and constituants.

The information expected in outputs for theses use cases are the one listed in section 1.3. As the use of the plateform is still recent and its features are still under heavy development, some of the use cases listed above are not entirely feasible at the time of writing but are currently desirable outcomes for ANSSI's CTI team.

4 Appendixes

4.1 Appendix 1: implemented datasets

In order to facilitate integration and ensure consistency in the global database, several datasets are implemented in the platform:

- the MITRE ATT&CK and pre-ATT&CK matrices for TTPs;
- sectors of activity;
- regions and countries in the world;
- the CVE list maintained by MITRE.

These datasets can be modified if they contain errors or are missing elements. An issue to change the dataset should be opened directly in the GitHub repository of the project to ensure the consistency of the dataset. New datasets can be added if necessary. Connectors also allow for the importation of specific datasets, such as the list of malwares brought by the Malpedia connector.

4.2 Appendix 2: definition of confidence levels

The confidence levels implemented in OpenCTI are based on a condensed version of the Admiralty code. It is a system designed for evaluating information, based on both the evaluation of the information itself and on the evaluation of the source. This system is commonly implemented for instance in NATO country members. The following table presents for each confidence level the equivalent levels in the Admiralty system.

Name	Information accuracy	Source accuracy
1 - Low	6 - Truth cannot be judged: No basis exists for evaluating the validity of the information | 5 - Improbable: Not confirmed; not logical in itself; contradicted by other information on the subject	F - Reliability cannot be judged: No basis exists for evaluating the reliability of the source | E - Unreliable: Lacking in authenticity, trustworthiness, and competency; history of invalid information
2 - Moderate	4 - Doubtful: Not confirmed; possible but not logical; no other information on the subject	D- Not usually reliable: Significant doubt about authenticity, trustworthiness, or competency but has provided valid information in the past
3 - Good	3 - Possibly True: Not confirmed; reasonably logical in itself; agrees with some other information on the subject | 2 - Probably True: Not confirmed; logical in itself; consistent with other information on the subject	C - Fairly reliable: Doubt of authenticity, trustworthiness, or competency but has provided valid information in the past | B - GUsually reliable: Minor doubt about authenticity, trustworthiness, or competency; has a history of valid information most of the time
4 - Strong	1 - Confirmed by other sources: Confirmed by other independent sources; logical in itself; Consistent with other information on the subject	A - Completely reliable: No doubt of authenticity, trustworthiness, or competency; has a history of complete reliability

4.3 Appendix 3: useful external resources

The following resources can be useful to understand and properly implement this doctrine:

- the general documentation on how to navigate and use OpenCTI is available at the following adress: `https://www.notion.so/OpenCTI-Public-Knowledge-Base-d411e5e477734c59887dad3649f20518`;
- the repository for the project is available here: `https://github.com/OpenCTI-Platform/opencti`;

- the MITRE ATT&CK and pre-ATT&CK matrices are available on the official website: https://attack.mitre.org/matrices/enterprise/ and can be useful to find the adequate TTPs;
- the official Oasis website for STIX: https://oasis-open.github.io/cti-documentation/stix/intro. Currently, OpenCTI uses STIX 2.1 format.

1.0 - 07/01/2021

AGENCE NATIONALE DE LA SÉCURITÉ DES SYSTÈMES D'INFORMATION

ANSSI - 51 boulevard de la Tour-Maubourg, 75700 PARIS 07 SP
www.cert.ssi.gouv.fr / cert-fr.cossi@ssi.gouv.fr

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Premier ministre
SGDSN