Riskaware

24th April 2017

# RA-OSINT

## *Riskaware Open-Source Intelligence*

### Overview

The RA-OSINT tool is designed to support the rapid acquisition and analysis of open-source intelligence for a variety of applications including Cyber Vulnerability Investigations (CVI) OSINT, threat intelligence, investigative journalism and police/intelligence investigations. It is based on a prototype developed in response to a Dstl Centre for Defence Enterprise (CDE) competition on Information Sense-Making. The tool has since undergone significant development, incorporating a flexible web search and ingest capability, together with more advanced analysis options and improved visualisation support. The aim has been to provide a highly user-interactive work-flow, supported by rich visual analytics, that allows the user to look deep into the data, rapidly bringing to the forefront the most important and relevant information, highlighting new details and then drilling down deeper into the relevant information to uncover new conclusions.

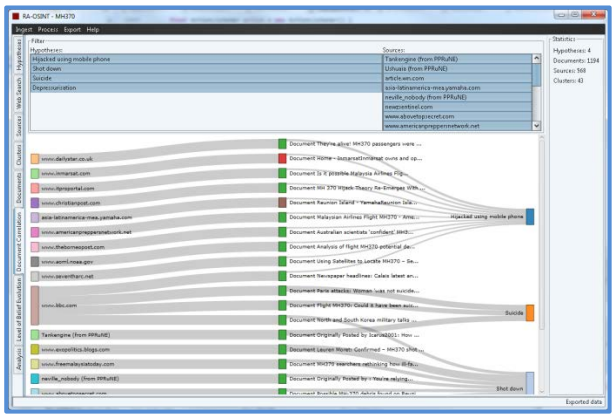

*Figure 1: RA-OSINT Sankey diagram, showing source and document correlation to hypotheses*

While it has been developed primarily with open-source intelligence in mind, it is equally capable when operating with closed-source data, for example archives of documents (seized, leaked, etc.) and material retrieved from the deep or dark web. The tool includes a custom ingestion framework that can perform bulk ingest from any well-structured web source, including forums, social media, marketplaces, etc.

### Typical Applications

- OSINT during Cyber Vulnerability Investigations (CVI);
- Threat intelligence, including searching for signs of activity by cyber criminals or inadvertent leakage of information;
- Investigative journalism, particularly when analysing large volumes of communication or documentation;
- Police and intelligence investigations;
- Ongoing monitoring of known "threat hypotheses" against latest news stories and social media posts;
- Identification of emerging concerns within previously gathered intelligence and feeds.

### Technical Basis

RA-OSINT has been developed as a flexible framework architecture into which we have integrated a number of industry-leading open-source software libraries. It harnesses recent developments within the open-source software community, including:

- Open-source tools for processing structured and unstructured information, particularly the ElasticSearch search and analytics engine and the Apache Carrot clustering and topic extraction engine[2];
- OrientDB connected graph database technology, that allows complex, highly connected datasets to be created;
- Powerful Visual Analytics: All the analysis options are supported with dynamic, user interactive visualisations, mostly based on the Data Driven Documents (D3) web visualisation library, to allow analysts to assimilate and interpret complex information.

### Key Capabilities

The tool includes a number of key elements that work together to provide the complete capability:

**Ingestion**:

To perform the important data capture task, RA-OSINT provides various tools, including both standard ingesters for well-known streams and a custom ingestion framework that allows the user to rapidly design ingesters for any regular, well-structured sites, for example forums or social media streams.

Since web search engines form an irreplaceable element of typical OSINT activity, the tool allows

the user to ingest selectively from web search outputs produced by key search engines. It also allows ingestion of a range of document types from local or network file stores, allowing closed-source content to be analysed.

**Analysis:**

The user can perform a variety of analysis activities on the invested data, and use the powerful visualisation features to review the results and drive further ingestion and analysis. Analysis can include:

- **Hypothesis analysis,** where the analyst enters hypotheses about the data and reviews the documents that most strongly link to the hypotheses being considered (see Figure 1);
- **Document relationship analysis**, where the documents relevant to current hypotheses are cross-compared to uncover associations and groupings, allowing the key documents to be identified and reviewed. This helps the analyst to contextualise the documents and identify important new ones (see Figure 2);

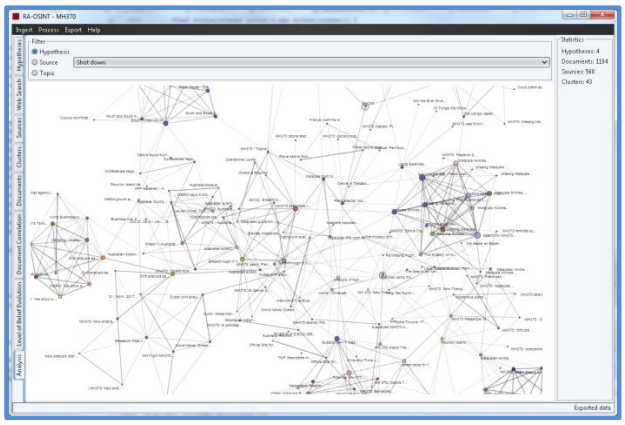

*Figure 2: RA-OSINT Force-directed graph of document relationships, for uncovering thematic linkage*

- **Topic extraction,** where relevant documents are mined for key phrases that are common amongst several documents. This can allow the discovery of new hypotheses or emerging concerns within the data (see Figure 3);


*Figure 3: Interoperability with Carrot² Workbench for topic analysis*

- **Document source and provenance analysis,** where relevant documents are traced back to source to allow analysis of provenance and reliability, and to identify "information incest" where multiple documents trace back to a single **source** (see Figure 1).
- **Timeline analysis,** where key documents are plotted on a timeline to analyse the evidence chronology that leads up to the current understanding (see Figure 4).

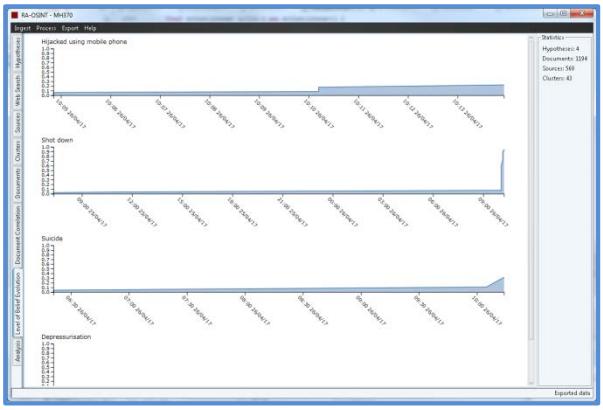

*Figure 4: Level of belief evolution timeline for key hypotheses*

**Logging:**

RA-OSINT provides full logging of user actions, together with archiving of all ingested material, providing traceability of material back to its source, together with post-review of analysis decisions. Raw and ingested data is not modified during the analysis process.

## Availability

The RA-OSINT single-user desktop tool is now available in the UK for use on selected projects, and is supported by Riskaware OSINT technical consultancy team. Further developments of the RA-OSINT core tool are expected in the future, including a browser-based version for team use.

**Contact:** robert.gordon@riskaware.co.uk

+44 (0) 117 9330523

**Riskaware Ltd**

Colston Tower, Colston Street,

Bristol BS1 4XE, UK

Riskaware

capability through technology