29th July 2016

Using a machine learning approach to improve the intelligence analysis process

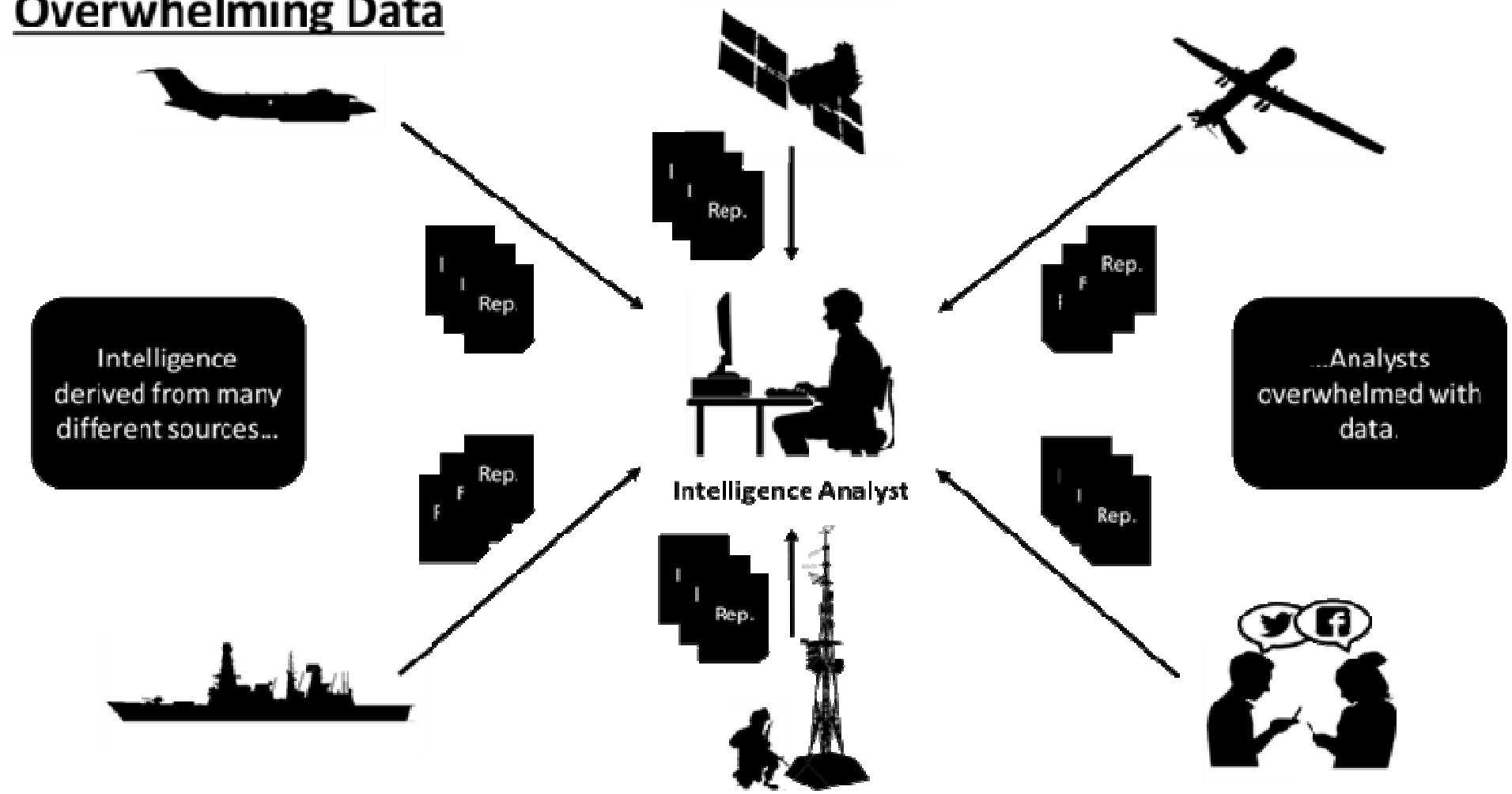
Tom Baldwin

Tel: 023 9225 9930

tombaldwin@polarisconsulting.co.uk

Background 1

Overwhelming Data

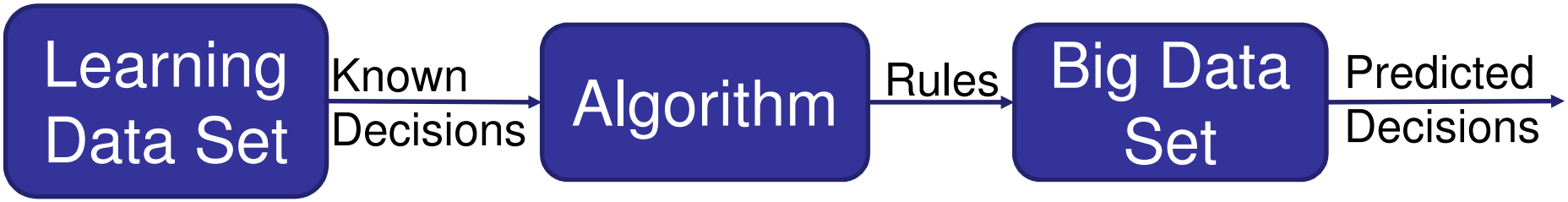


Background 2

- Enhance intelligence analysis
- Reduce analyst burden
- **Phase 1 – proof-of-principle**
  - Jan – June 2016
- Phase 2 – real-world testing
  - Oct 2016 – Sep 2017
  - Under consideration

Our Idea

- Decision support tool
  - Machine Learning

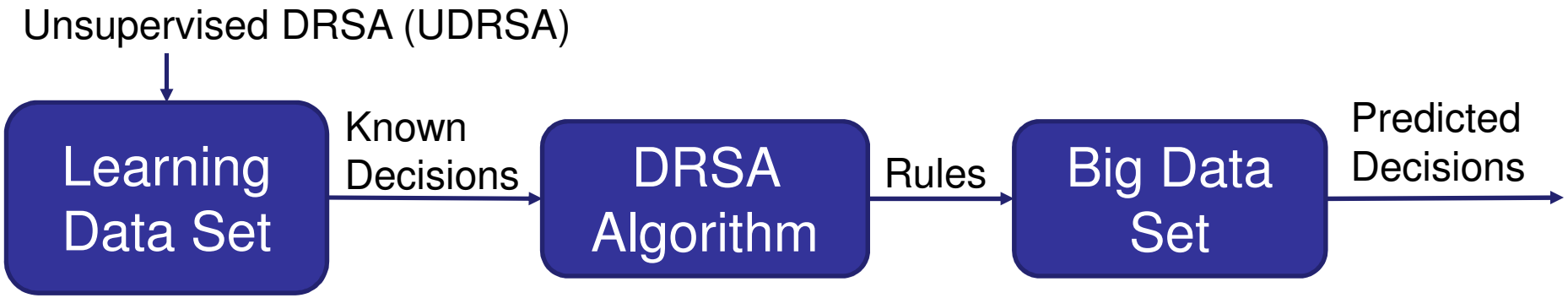


- Intelligence Analysis Application
  - Learn how analysts prioritise reports
  - Capture underlying rules e.g. key attributes
  - Dynamically prioritise new reports

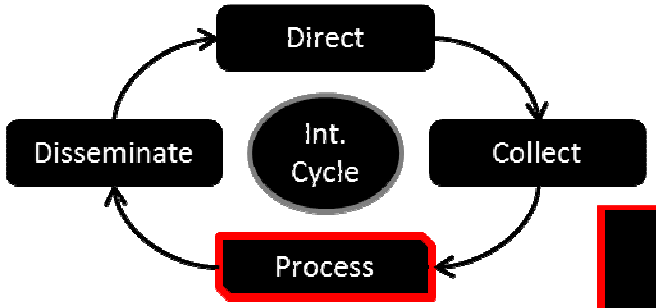
Created by Gregor Cresnar from Noun Project

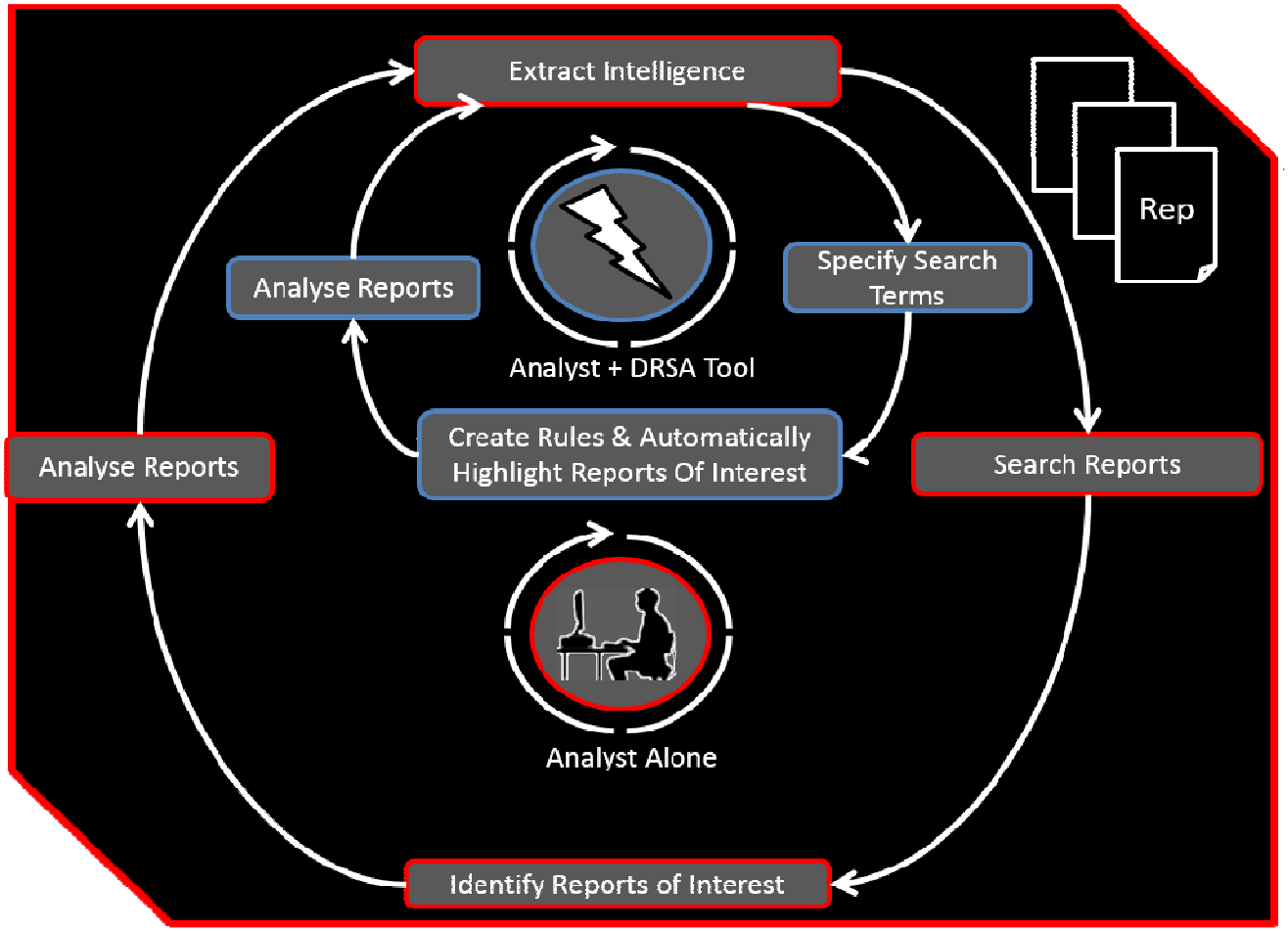
The Approach

- Dominance-based Rough Set Approach (DRSA)
  - Dominance relation replaces indiscernibility relation
  - Preference ordered attributes & decision classes
- Advantages
  - Multiple data formats
  - Addresses gaps / inconsistencies in data
  - Simple IF-THEN rules
  - Efficient & effective vs. other methods



Application 1





Application 2

1

UDRSA applied to all reports to identify representative learning set

Report No.	Date	Time	Location	Person of interest	⋮	Interest level
1	⋮	⋮	⋮	⋮	⋮	?
2	⋮	⋮	⋮	⋮	⋮	?
3	*⋮*	*⋮*	*Market*	*⋮*	*⋮*	*?*
4	⋮	⋮	⋮	⋮	⋮	?
5	*⋮*	*⋮*	*⋮*	*⋮*	*Apple*	*?*
6	⋮	⋮	⋮	⋮	⋮	?
7	⋮	⋮	⋮	⋮	⋮	?
8	⋮	⋮	⋮	⋮	⋮	?
9	*⋮*	*⋮*	*Market*	*⋮*	*Pear*	*?*
10	⋮	⋮	⋮	⋮	⋮	?
⋮	⋮	⋮	⋮	⋮	⋮	?
45	*⋮*	*⋮*	*Market*	*⋮*	*⋮*	*?*
⋮	⋮	⋮	⋮	⋮	⋮	⋮
99	*⋮*	*⋮*	*⋮*	*⋮*	*Melon*	*?*
100	⋮	⋮	⋮	⋮	⋮	?

2

Analyst specifies interest level to reports within learning set

Learning Set

Report No.	Date	Time	Location	Person of interest	⋮	Interest level
3	*⋮*	*⋮*	*Market*	*⋮*	*⋮*	*1*
5	*⋮*	*⋮*	*⋮*	*⋮*	*Apple*	*3*
9	*⋮*	*⋮*	*Market*	*⋮*	*Pear*	*5*
45	*⋮*	*⋮*	*Market*	*⋮*	*⋮*	*8*
99	*⋮*	*⋮*	*⋮*	*⋮*	*Melon*	*9*

Main Dataset

Report No.	Date	Time	Location	Person of interest	⋮	Interest level
1	⋮	⋮	⋮	⋮	⋮	?
2	⋮	⋮	⋮	⋮	⋮	?
4	⋮	⋮	⋮	⋮	⋮	?
6	⋮	⋮	⋮	⋮	⋮	?
⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	⋮	⋮	⋮	⋮	⋮	?

Application 3

3 **DRSA** applied to learning set to generate IF-THEN rules

Report No.	Date	Time	Location	Person of interest	⋮	Interest level
3	⋮	⋮	Market	⋮	⋮	1
5	⋮	⋮	⋮	⋮	Apple	3
9	⋮	⋮	⋮	⋮	Melon	5
45	⋮	⋮	Market	⋮	⋮	8
99	⋮	⋮	Market	⋮	Melon	9

Rules

1. *If* Market *then* Int. Level ≥ 1
2. *If* Apple *then* Int. Level ≥ 3
3. *If* Melon *then* Int. Level ≥ 5
4. *If* Market & Melon *then* Int. Level ≥ 9

4 Rules applied to unread reports

5 High interest reports brought to analyst's attention, then scored & added to learning set

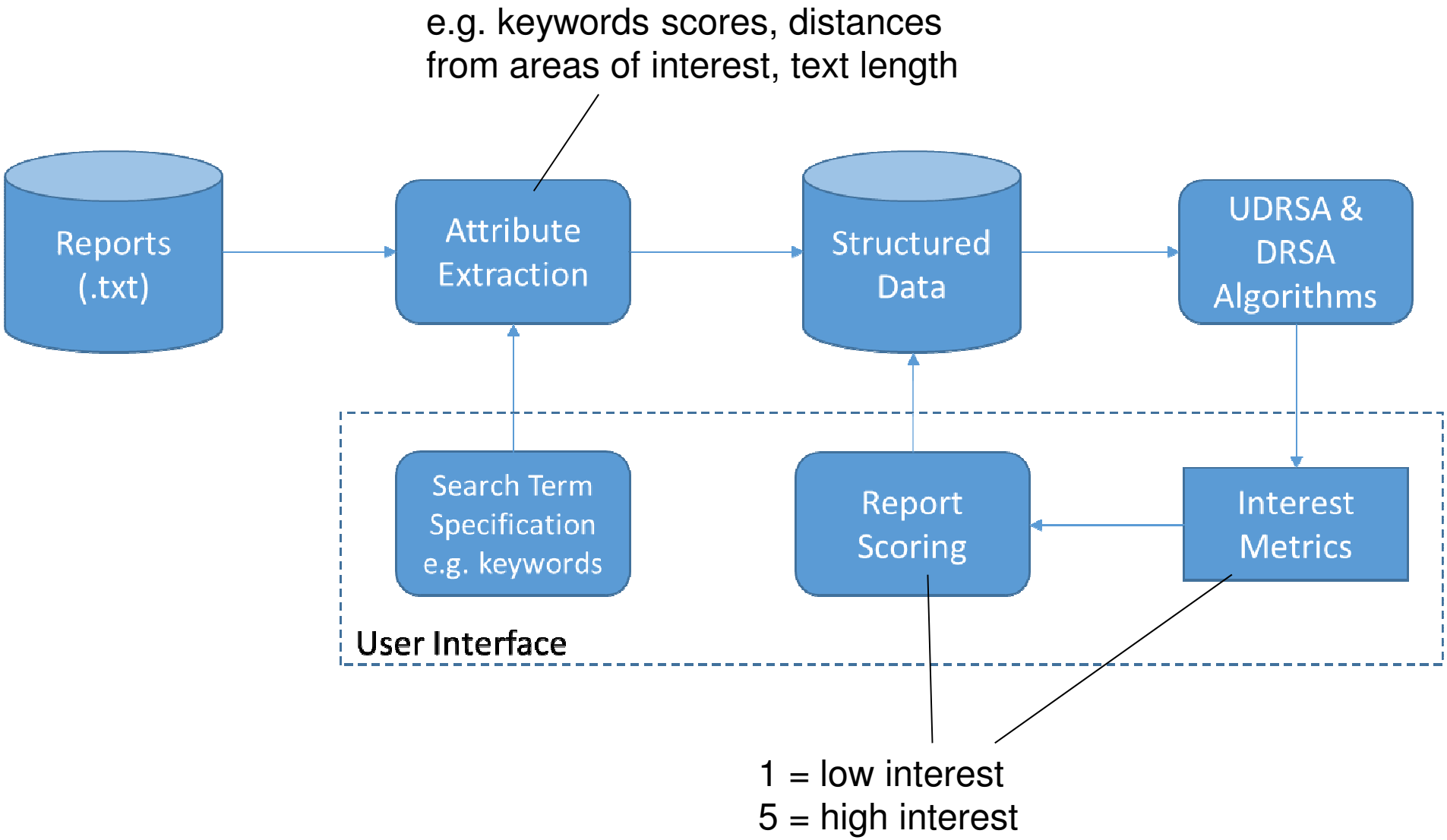
Report No.	Date	Time	Location	Person of interest	⋮	Interest level	DRSA Predicted Interest
1	⋮	⋮	⋮	⋮	⋮	4	4
2	⋮	⋮	⋮	⋮	⋮	9	6
4	⋮	⋮	⋮	⋮	⋮	7	4
6	⋮	⋮	⋮	⋮	⋮	5	5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	⋮	⋮	⋮	⋮	⋮	8	4

Main Dataset

Report No.	Date	Time	Location	Person of interest	⋮	Interest level	DRSA Predicted Interest
1	⋮	⋮	⋮	⋮	⋮	?	4
2	⋮	⋮	⋮	⋮	⋮	?	6
4	⋮	⋮	⋮	⋮	⋮	?	4
6	⋮	⋮	⋮	⋮	⋮	?	5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	⋮	⋮	⋮	⋮	⋮	?	3

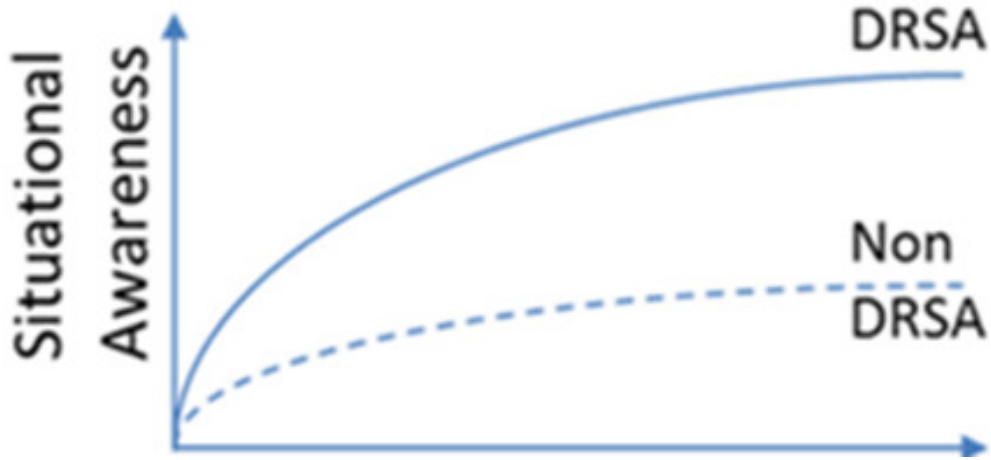
Main Dataset

DRSA Tool Design



Experiment Hypothesis

- DRSA tool enables analysts
  - to quickly identify & prioritise high interest reports
  - ignore less important reports
  - extract more useful intelligence, more quickly & gain better situational awareness (SA)



Experiment Bounding

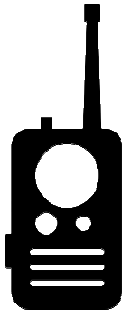
- Future Electromagnetic Environment
  - Very complex
  - High volumes of data

Created by Ruben Vin from Noun Project

- Strategic Electronic Warfare (EW)
  - Millions of intercepts per day
  - Up to 300 fields

Created by Creative Stall from Noun Project

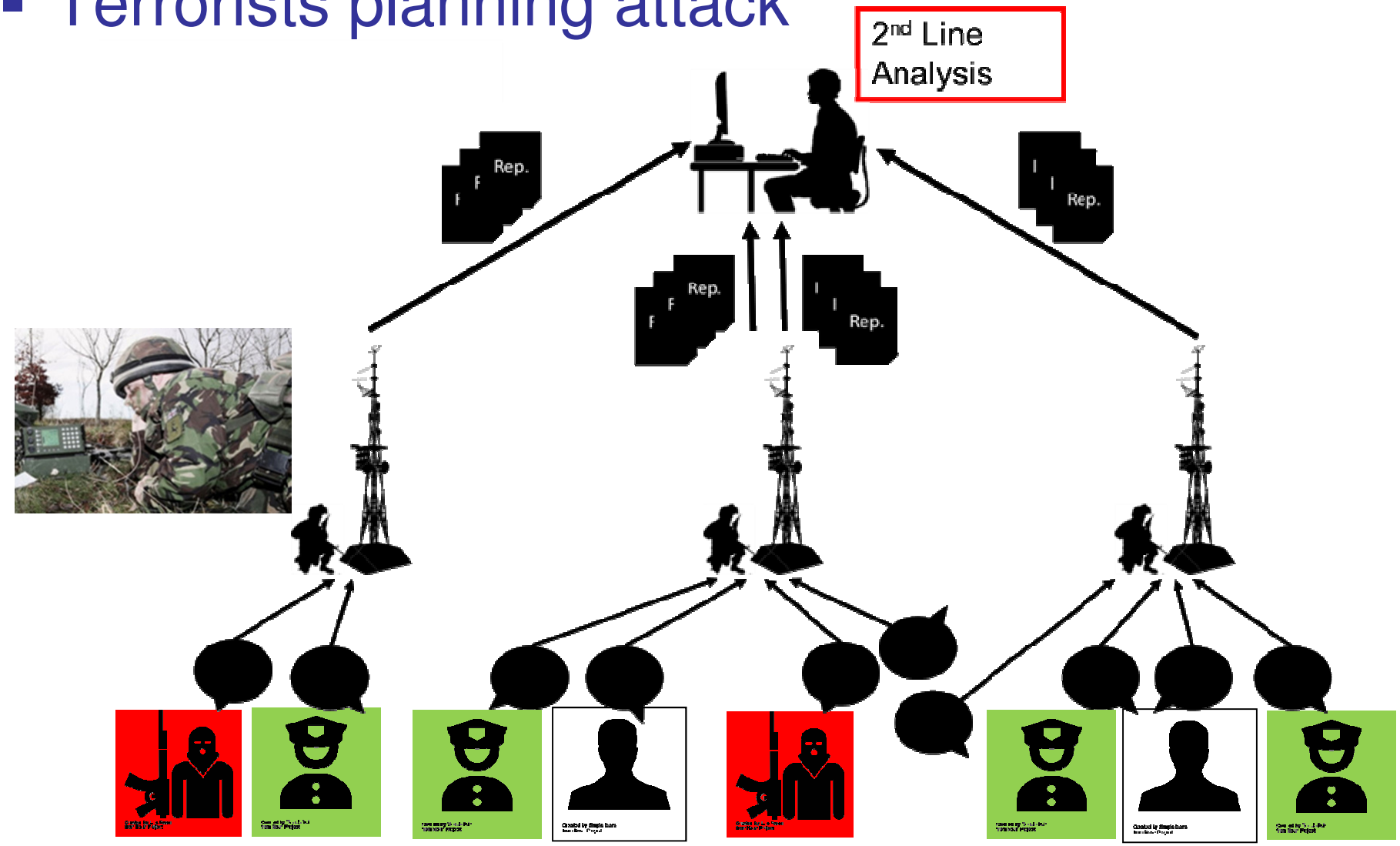
- Tactical Signals Intelligence (SIGINT)
  - Collection / Analysis balance
  - NATO collaboration initiatives



Created by Linda Yuki Nakanishi from Noun Project

Experiment Scenario 1

- Terrorists planning attack



Experiment Scenario 2

- 450+ reports

Day Time Group

Platform – intercept source *(out of scope)*

Frequency e.g. VHF/UHF

To/From *(UIM/F = Unidentified Male/Female)*

Transmission Locations *(out of scope)*

Gist = Summary of what was discussed

```
DTG: 01 0605
Platform: LEWT 4
Freq: 147.955
To: UIM
From: BIG BIRD
Loc: 34548934

Gist:

UIM             BIG BIRD, ARE YOU THERE YET?

BIG BIRD        YES I AM.

UIM             OK, GREAT. DON'T FORGET TO USE THE RHINO LIKE I TRAINED YOU.

BIG BIRD        I WILL. IF IT DOES AS MUCH DAMAGE WE'LL ALL BE VERY HAPPY.

UIM             HA GOOD LUCK BOB, MY FRIEND. OUT.

Op Comment:     BIG BIRD IS POSSIBLE CODE NAME FOR BOB. RHINO IS POSSIBLE COVER TERM FOR WEAPON.
```

Operator Comments *(Initial 1st Line Analysis)*
Code names e.g. Big Bird = Bob
Cover terms e.g. Rhino = Weapon

Experiment Set Up

- 11 participants
  - 6 x DRSA Tool
  - 5 x non-DRSA Tool
- Training
- Intelligence Briefing
- 3 x 45 minutes exercises
  - Review & score reports
  - Extract intelligence
  - INTSUM (v Ground Truth)
  - User Survey

DRSA Tool 1

Report Summary Attributes

User Defined Interest Score

Predicted Interest Scores

0 / 34 Reports Scored **00:00:13**

Mission | Reports

	File	DTG	Freq	To	From	Interest	DRSA	U-DRSA
1	DAY1_0029.txt	01 0534	136.136	UIM	Jimmy			
2	DAY1_0001.txt	01 0730	159.662	Ekotto	Dave			
3	DAY1_0002.txt	01 0800	159.662	James	Louis			
4	DAY1_0003.txt	01 0805	187.333	Naiomi	Faye			
5	DAY1_0004.txt	01 0807	144.453	James	Jeremy			
6	DAY1_0005.txt	01 0810	194.677	Samuel	Jenna			
7	DAY1_0030.txt	01 0825	136.136	Seth	UIM			
8	DAY1_0006.txt	01 0830	159.662	Lisa	Tina			
9	DAY1_0007.txt	01 0910	72.101	Gareth	Faisal			
10	DAY1_0008.txt	01 0920	116.878	Judy	Nick			
11	DAY1_0009.txt	01 0940	90.652	Mike	Lee			
12	DAY1_0022.txt	01 0942	102.438	Steve	Jamie			
13	DAY1_0010.txt	01 0950	108.672	Paul	Chris			
14	DAY1_0011.txt	01 1005	159.662	Kolo	Smith			
15	DAY1_0020.txt	01 1005	88.843	Steve	Wayne			
16	DAY1_0012.txt	01 1010	187.333	Heather	Julia			
17	DAY1_0013.txt	01 1015	124.396	Trent	Robert			
18	DAY1_0014.txt	01 1020	137.852	Nigel	Andy			
19	DAY1_0015.txt	01 1030	107.481	Kiki	UIM			
20	DAY1_0023.txt	01 1036		UIM	Jamie			
21	DAY1_0028.txt	01 1037	136.136	Seth	Jimmy			
22	DAY1_0016.txt	01 1100	176.231	Sibyelle	Timothy			
23	DAY1_0017.txt	01 1104	90.652	Mark	Robert			
24	DAY1_0018.txt	01 1115	98.123	UIM	UIF			

DRSA Tool 2

Report 79 - DAY1_0079.txt

To	From	at	on
ANYA	FRANK	01 1230	153.325

Gist:

UIF FRANK FRANK, YOU THERE?

FRANK YUP I CAN HEAR YOU CLEARLY.

UIF HOW DOES IT LOOK, ARE THE SPARROWS STILL THERE?

FRANK YES THEY KEEP COMING BACK. THERE ARE A LOT OF BEES HERE AS WELL.

People:		Places:		Keywords:	
	Add		Add		Add
MARA	Up	THIRTY JAGOWSTRA	Up	CLOVER	Up
LEVIATHAN	Down	HOTPOT	Down	SEEDS	Down
ULRICH	Remove	BRANDENBURG GATI	Remove	457	Remove
ANYA		CHARIOT		CARPET	
HANS		REICHSTAG		BEE	
ULRIKA		JAGERSTRASSE		RECCES	
SONYA				PEAS	
FRIEDA					

Interest: 5 4 3 2 1 Close

DRSA Tool 3

Report Summary Attributes

User Defined Interest Score

Predicted Interest Scores

DRS CDE

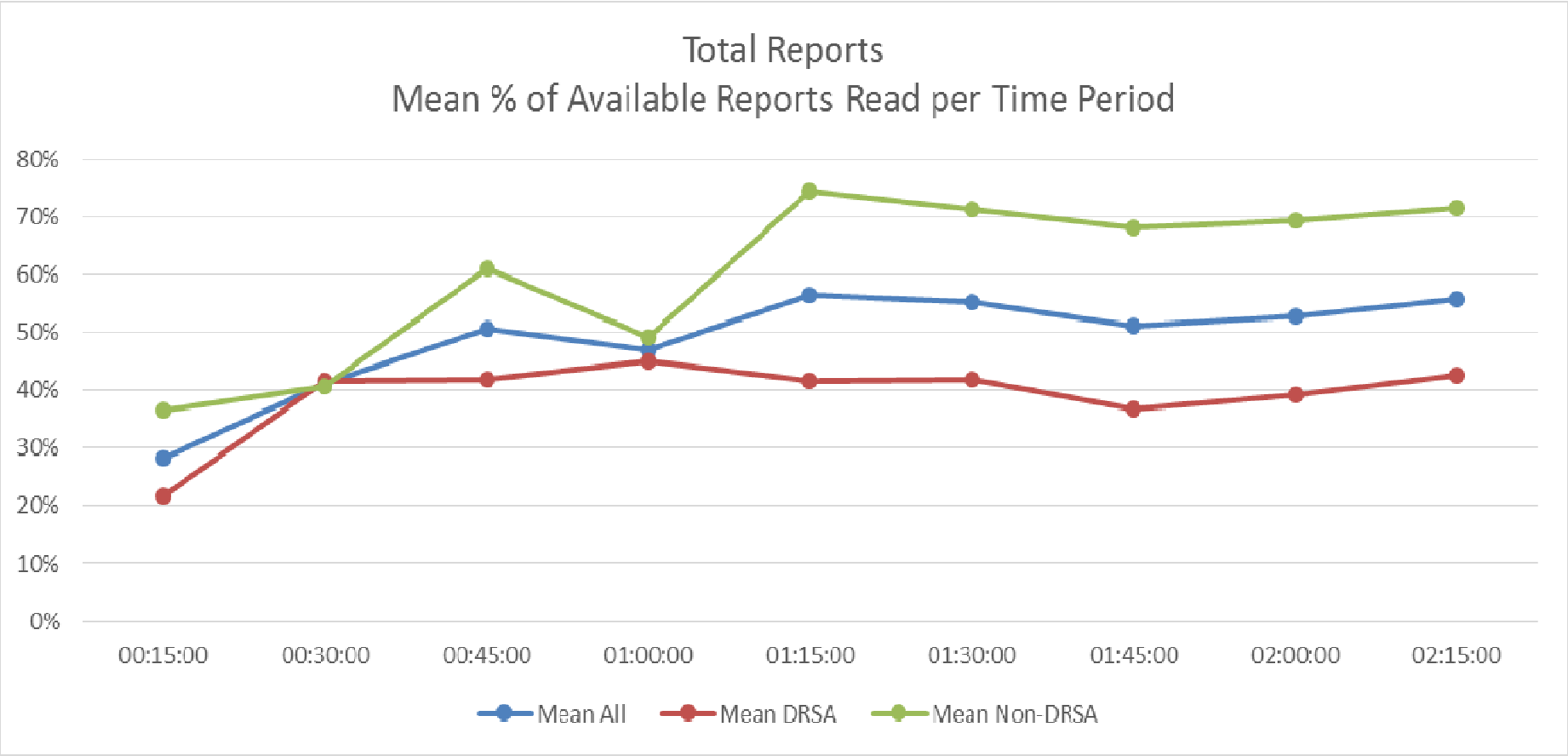
156 / 496 Reports Scored 02:18:1

Mission | Reports

	File	DTG	Freq	To	From	Interest	DRSA	U-DRSA
29	DAY1_0029.txt	01 0721	154.325	UIM1		1	1	5
30	DAY1_0030.txt	01 0730	153.325	FRANK	ANYA	4	4	5
31	DAY1_0031.txt	01 0733	147.955	UIM	COBALT		1	1
32	DAY1_0033.txt	01 0740	147.955	UIM	COBALT		1	1
33	DAY1_0032.txt	01 0745	153.325	ANYA	FRANK	4	4	3
34	DAY1_0034.txt	01 0753	147.955	UIM	COBALT		1	1
35	DAY1_0035.txt	01 0800	153.325	(FRANK)	ANYA	4	4	3
36	DAY1_0036.txt	01 0802	147.955	UIM	COBALT		1	1
37	DAY1_0037.txt	01 0805	147.955	UIM	COBALT		1	1
38	DAY1_0038.txt	01 0810	147.955	UIM	COBALT		1	1
39	DAY1_0039.txt	01 0813	142.835	UIM	PIPER-1		1	1
40	DAY1_0040.txt	01 0815	147.955	UIM	COBALT		1	1
41	DAY1_0041.txt	01 0817	154.325	UIF			1	1
42	DAY1_0042.txt	01 0822	147.955	UIM	COBALT		1	1
43	DAY1_0043.txt	01 0823	147.955	UIM	COBALT		1	1
44	DAY1_0044.txt	01 0824	147.955	UIM		1	1	1
45	DAY1_0045.txt	01 0825	145.475	ANYA	SONYA	4	4	5
46	DAY1_0046.txt	01 0837	147.95	UIM	COBALT	1	1	5
47	DAY1_0047.txt	01 0841	136.748	UIM		3	4	5

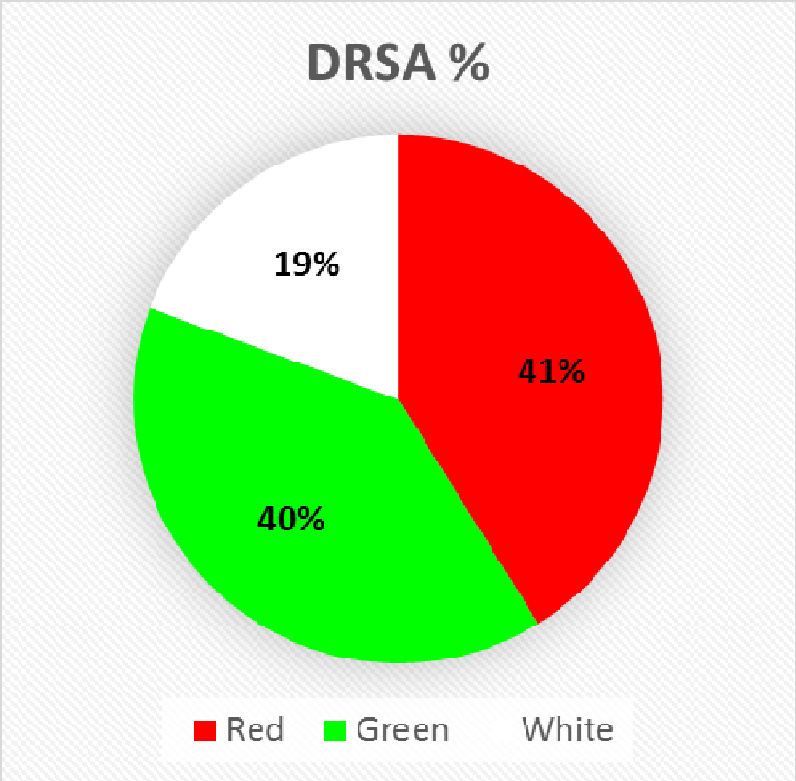
Results 1 – Reports Processed

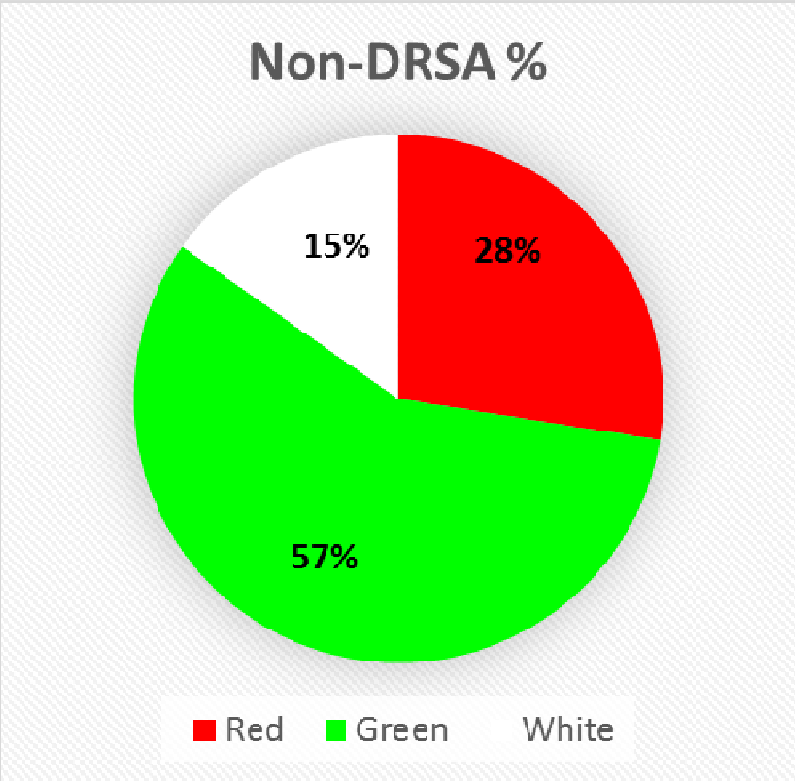
- DRSA analysts processed less reports



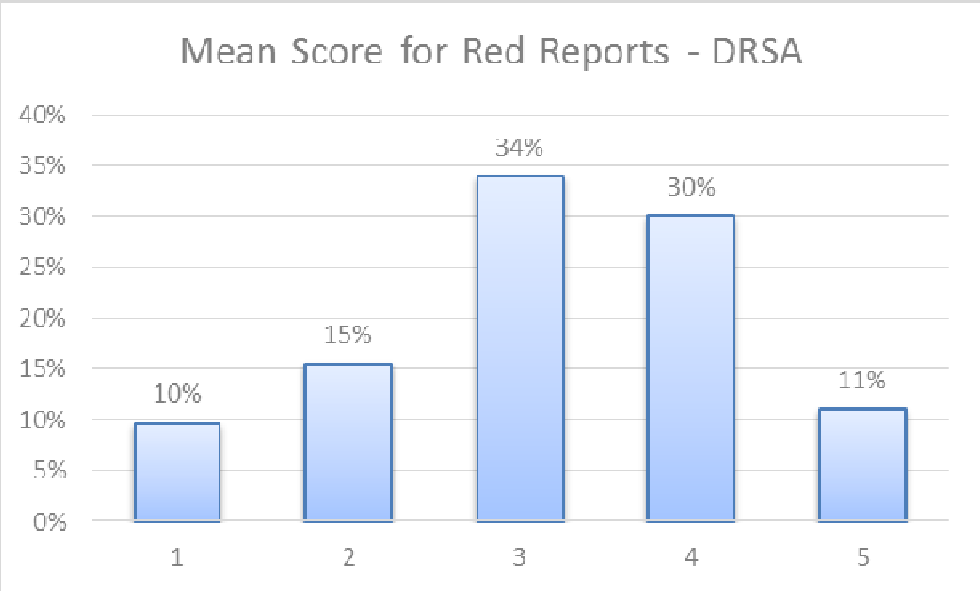
Results 2 – Reports Processed

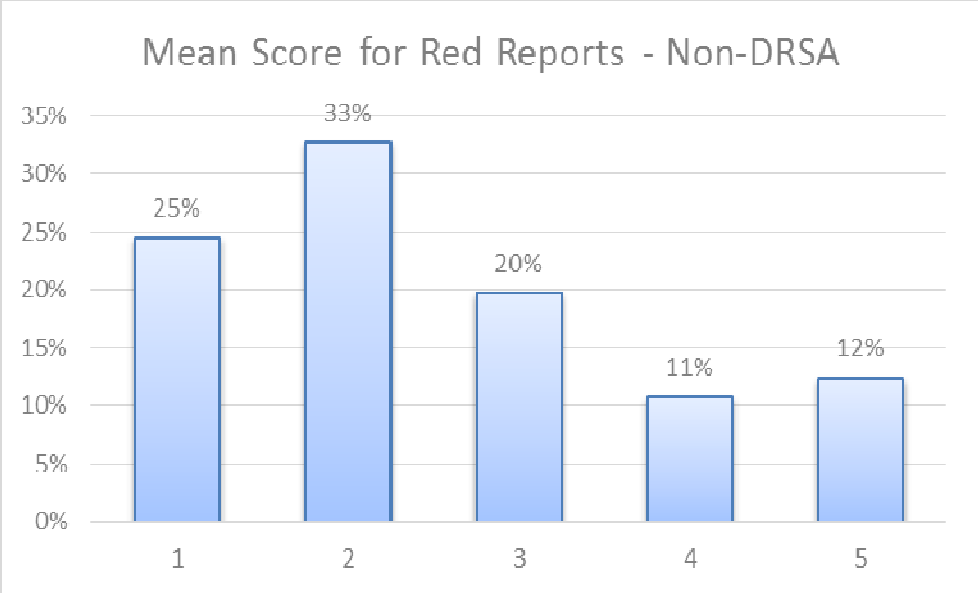
- DRSA analysts processed greater proportion of Red reports
  - Red = Terrorist, White = Civ., Green = Mil/Police





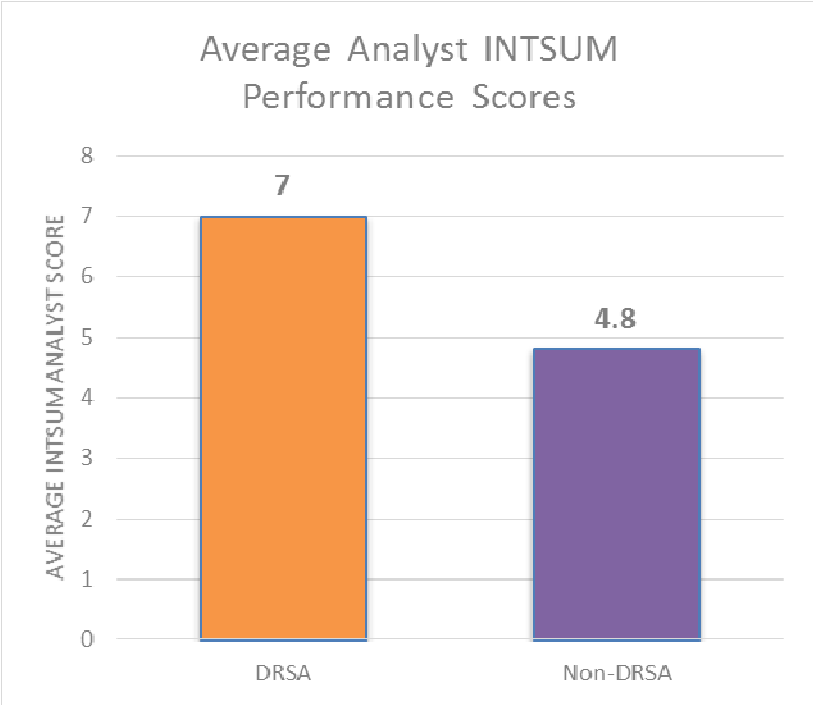
Results 3 – Red Report Scores

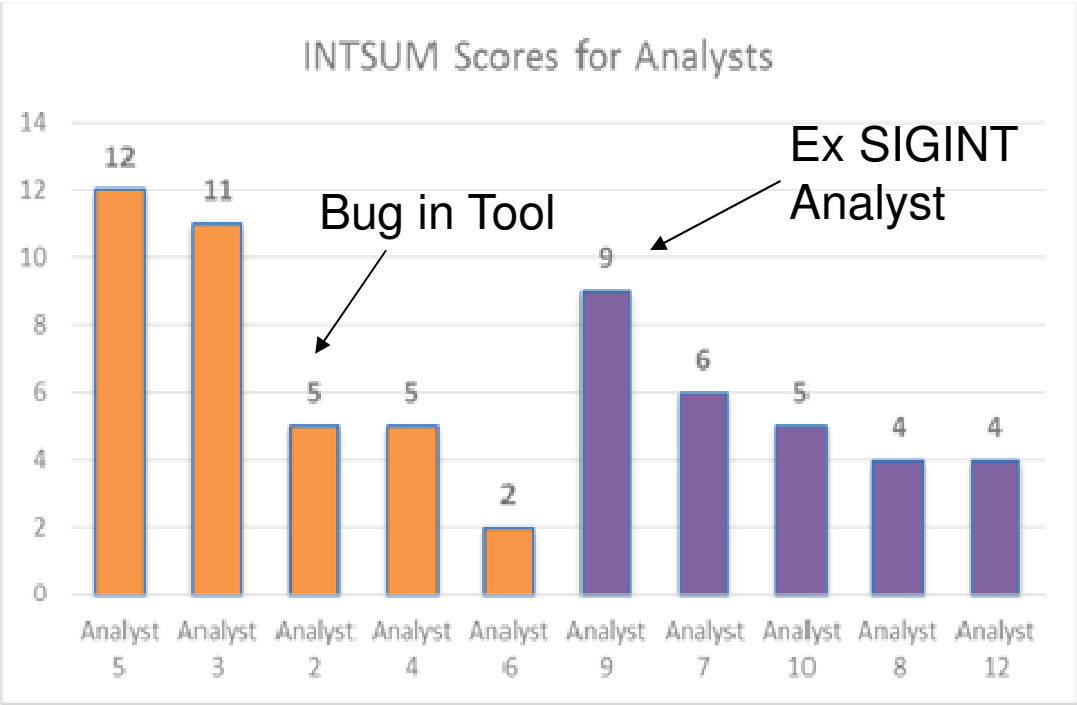




- DRSA analysts
  - scored Red (i.e. terrorist) reports highly
  - better able to correctly identify high interest reports

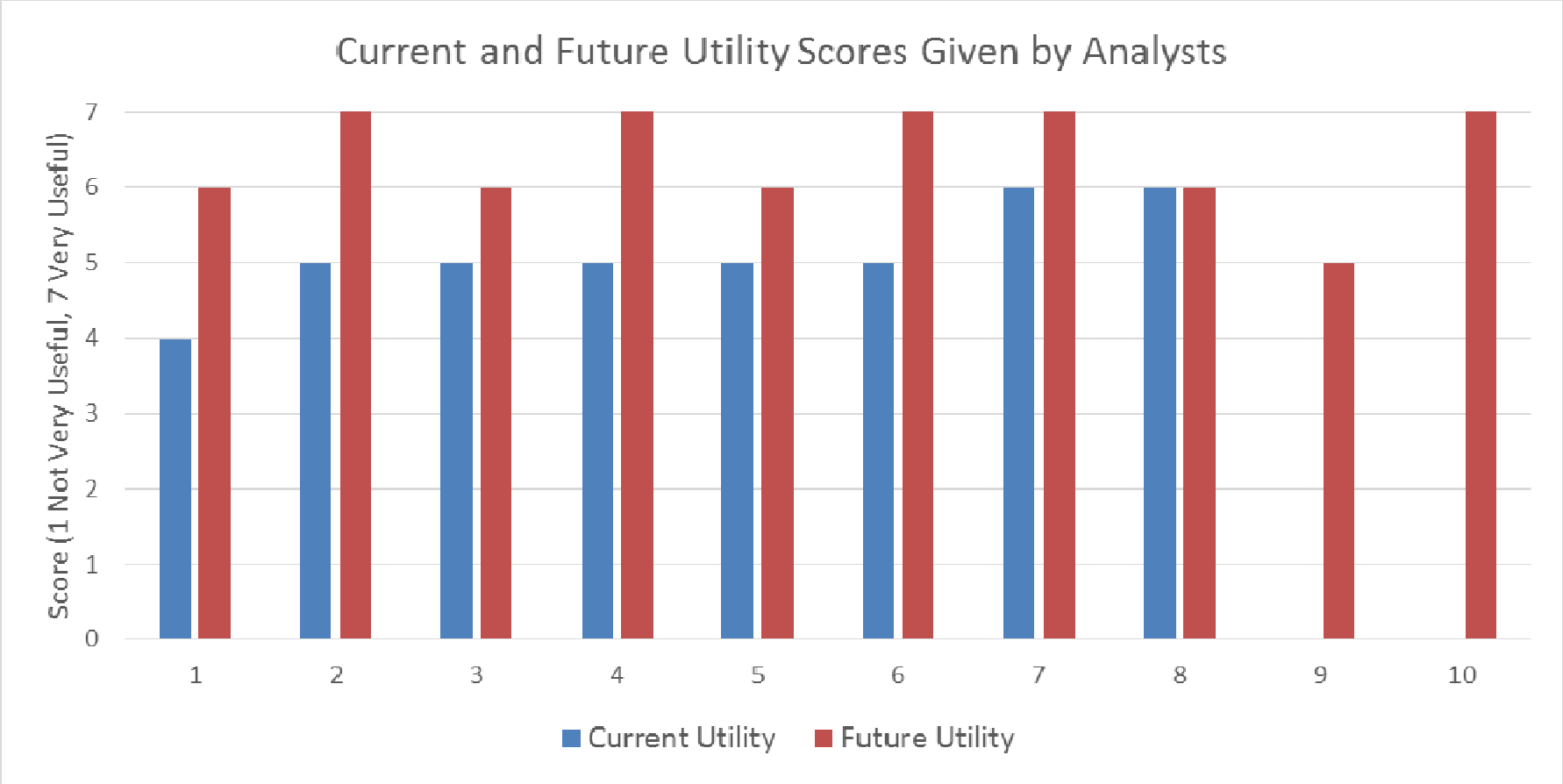
Results 4 – INTSUM Results





- DRSA analysts extracted more intelligence
- More a measure of analyst ability than DRSA?
  - Report Identification v Intelligence Extraction

Results 5 – User Feedback



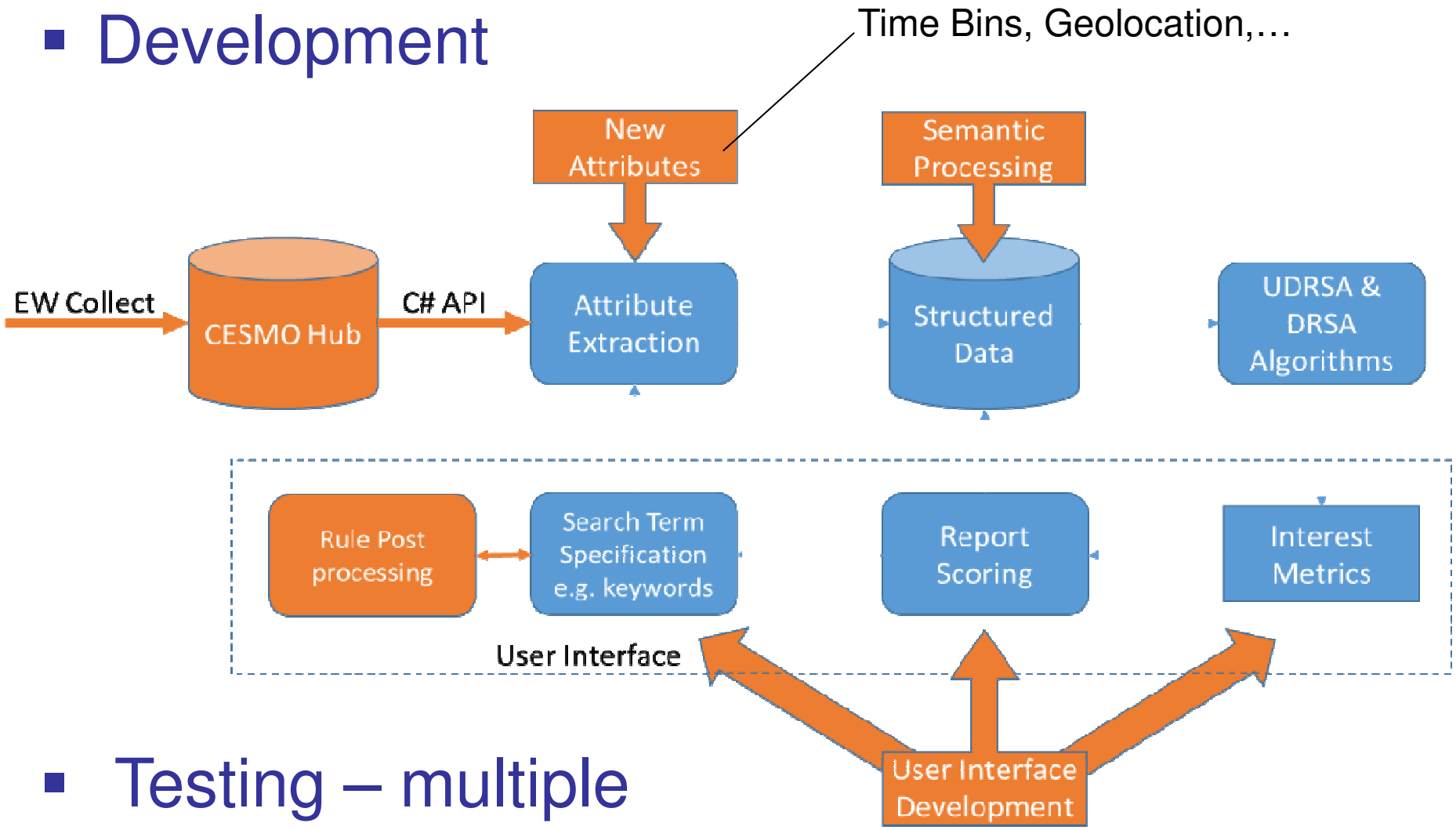
- Feedback on DRSA was very positive

Experiment Conclusions

- Evidence suggests DRSA enables analysts
  - to quickly identify & prioritise high interest reports & ignore less important reports
  - extract more useful intelligence, more quickly & gain better situational awareness
- Experiment Limitations
  - Small data set
  - Small sample size
  - Participants not intelligence analysts

Next Steps – CDE Phase 2?

- Development



- Testing – multiple intelligence analysts

Future Applications

- Intelligence Processing
  - EW/SIGINT analysis
  - Other Intelligence Domains
  - Other Government Departments
- Broader applications
  - Analyst training
  - Signature Recognition & Identification
  - Communication networks
- Non Defence

Summary

- Intelligence analysts faced with challenge of overwhelming data
  - Machine learning approaches can address this
- Developed & tested a DRSA tool
  - Shown potential to augment intelligence analysis processing
- Further work required to refine approach & its application to existing data & processes

Questions?

Tel: 023 9225 9930
tombaldwin@polarisconsulting.co.uk