Events Product Centre

GCHQ

events Agenda

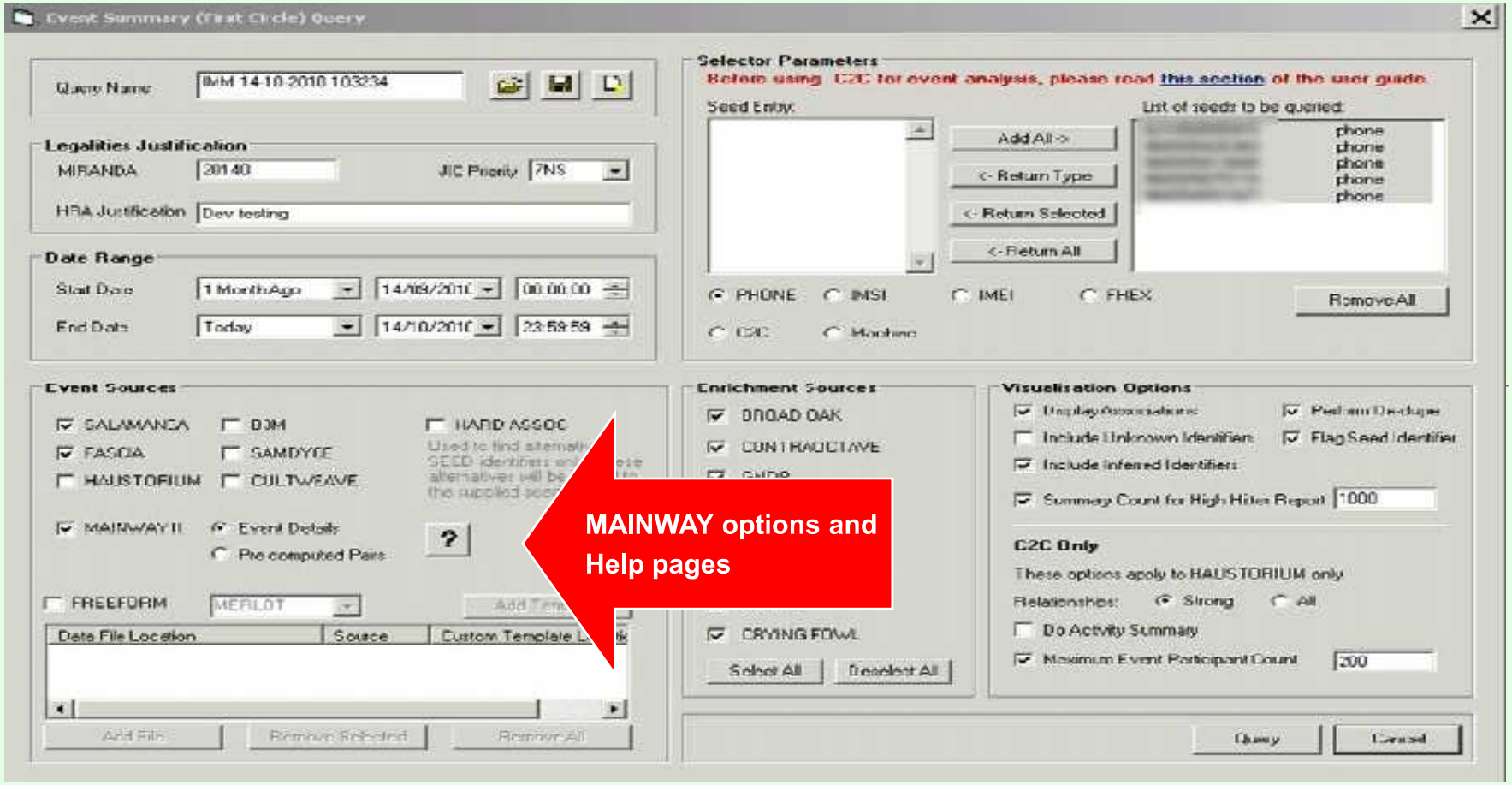
- Welcome
- Immingle
- Salamanca
- QFDs
- Guiding Light
- Questions

GCHQ

events IMMINGLE

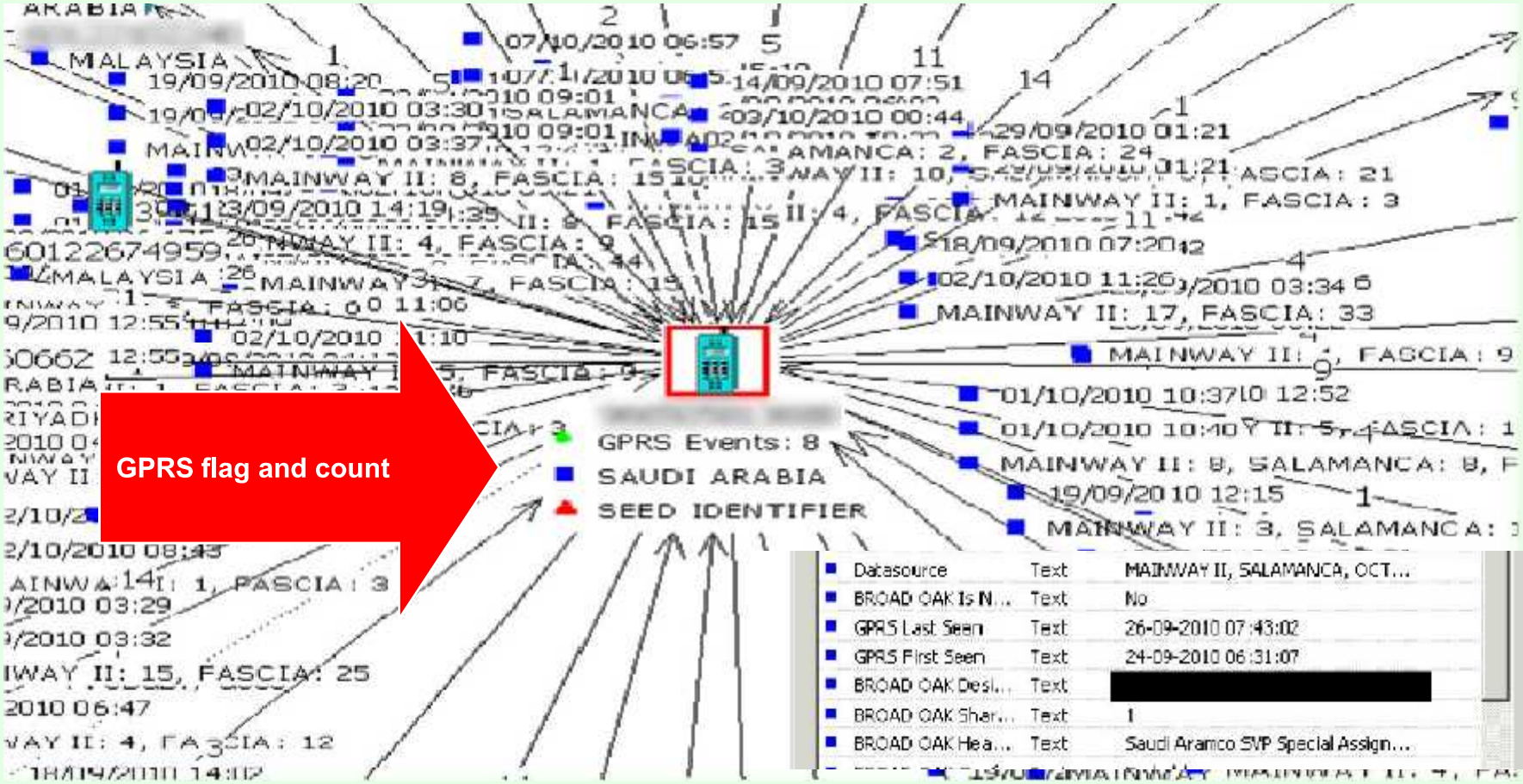
- Key changes July 2010 to present:
  - Inferred data from B3M now flagged
  - Updates to handle identifiers from HARD ASSOC and B3M correctly
  - MAINWAY: MSRNs now grouped and flagged in same way as SALAMANCA
  - MAINWAY: direct access to event details provided
  - GPRS flagging – THUGGEE rules applied to SALAMANCA events

events IMMINGLE



GCHQ

events IMMINGLE



GCHQ

events IMMINGLE

- What next?
  - FASCIA GPRS flagging
  - HAUSTORIUM decommissioning
  - Next Gen Contact Chaining trial….

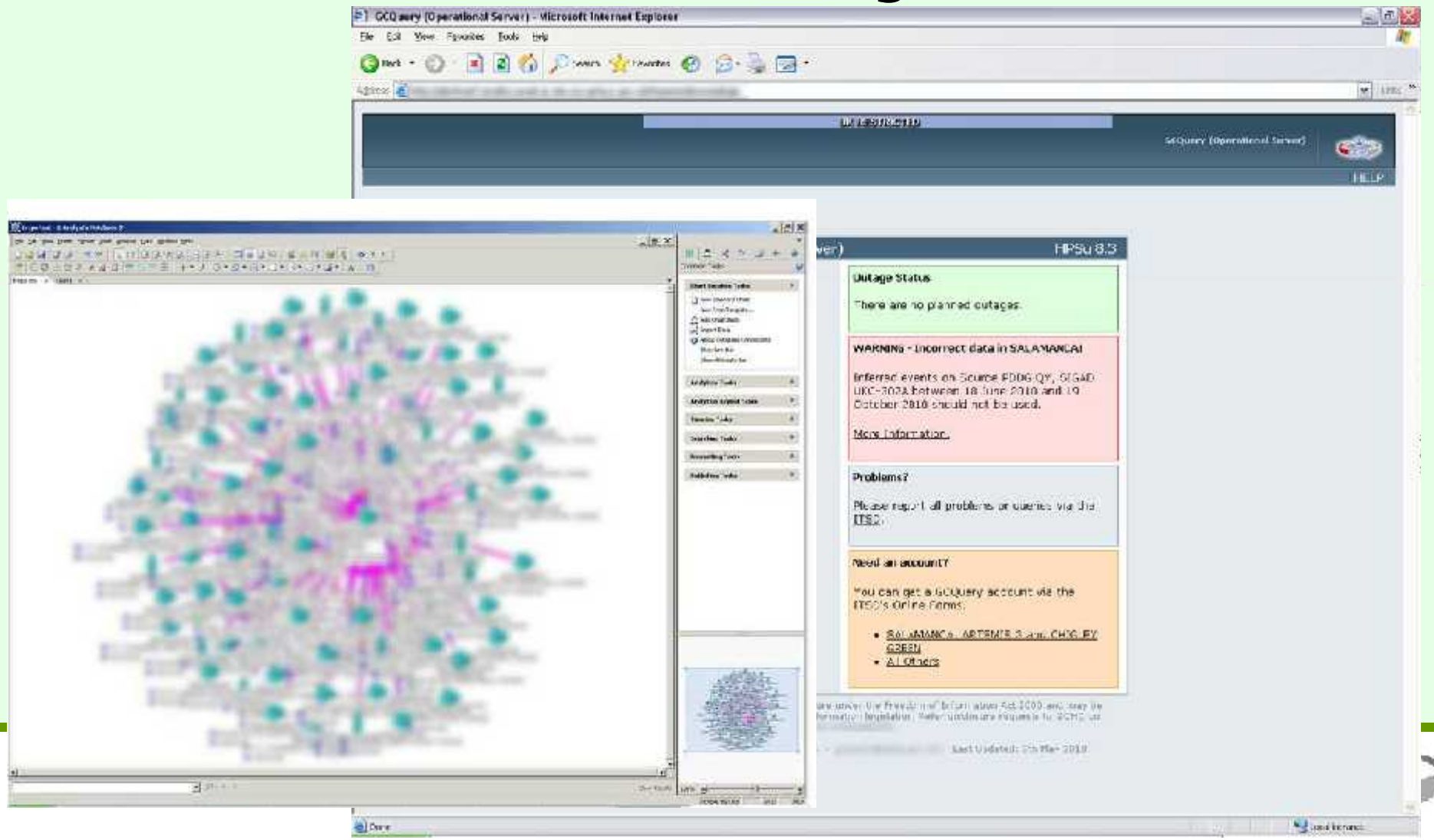
GCHQ

BRIO and SALAMANCA

- **Key changes since July:**
- NRT (Near Real Time) Storage = 3 days
- Extra feeds from TERRAINs at BUDE and SOUNDER
- 2nd Party usage of SALAMANCA: SHAREOWN replaces ESCHAR
- CallAnsweredState and CallEndState added to TERRAIN-SALAMANCA feed

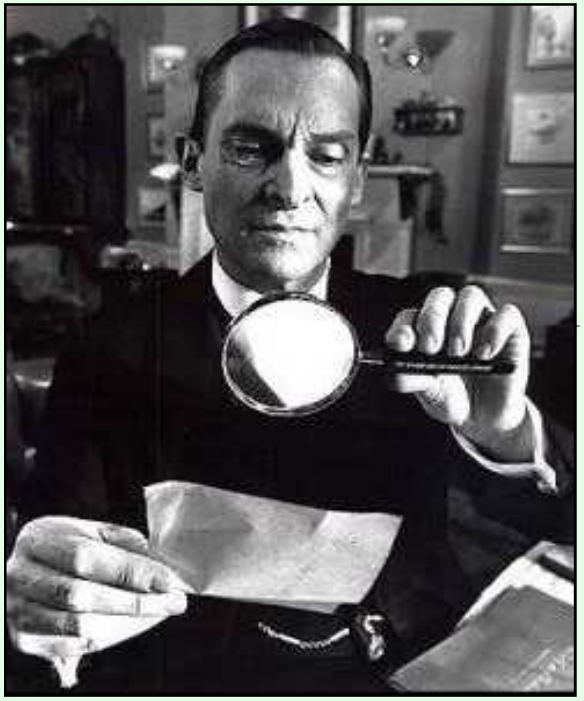
events BRIO/SALAMANCA (cont.)

- Pakistan NGN inferencing errors



events

QFDs



[redacted]

GCHQ

Scaling

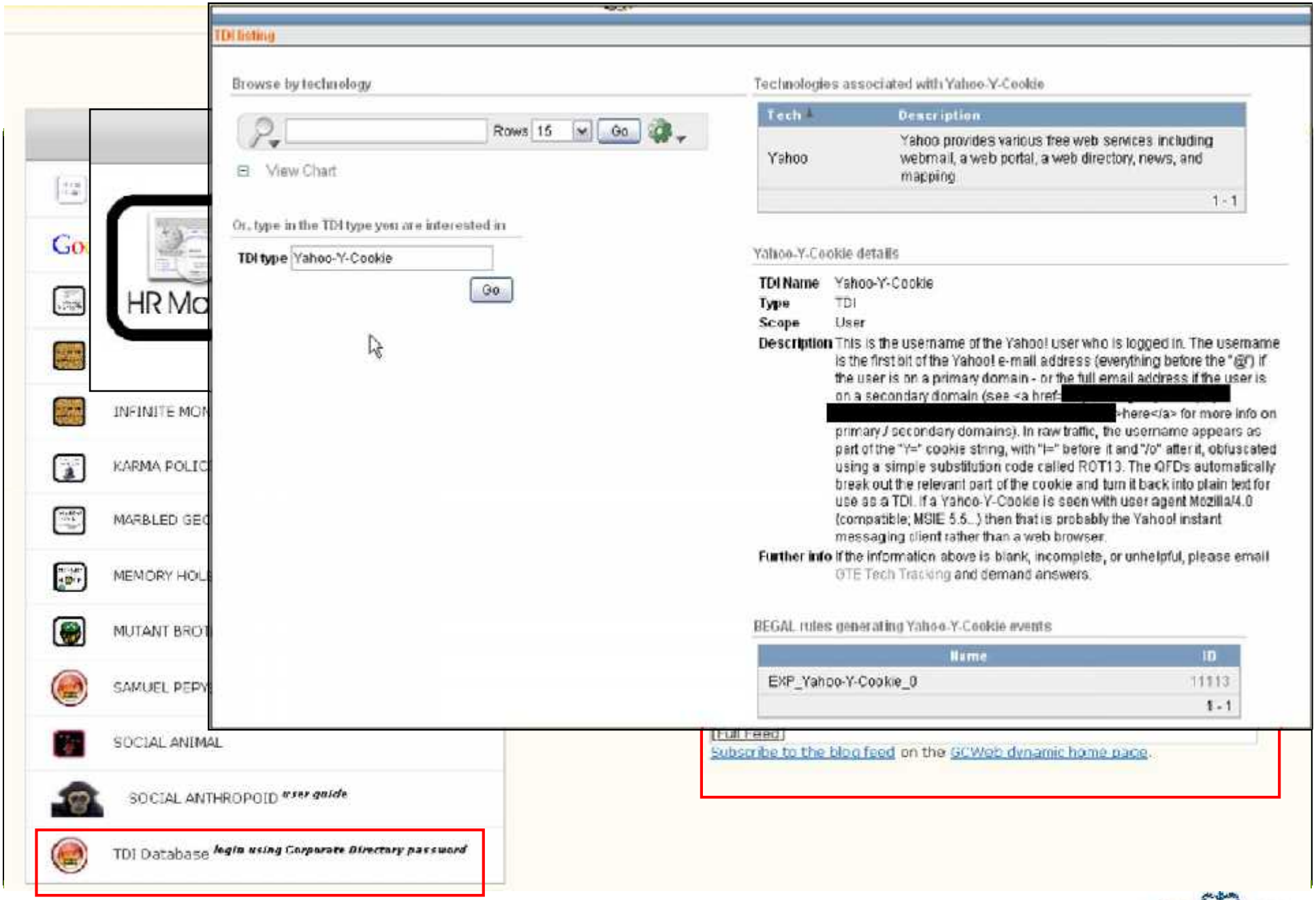
Current scale

- There are 100 unique bearers feeding the BzS tools.
- Consistently averaging over 30 billion events per-day into the input buffer.
  - MB is loading over 10.5 billion
  - 6 months data retention for MB = 1,890,000,000,000 records and requires 400 TB.
  - Total storage of over one petabyte.

Scaling

Future Scale

- Further 58 bearers by end of 2010
- An additional 40 bearers in Q1 2011.
  - MB will ingest over 20 billion events per day requiring one petabyte of storage.
  - Overall storage will increase to 2.5 petabytes.
- Scope scaling to 400 bearers.

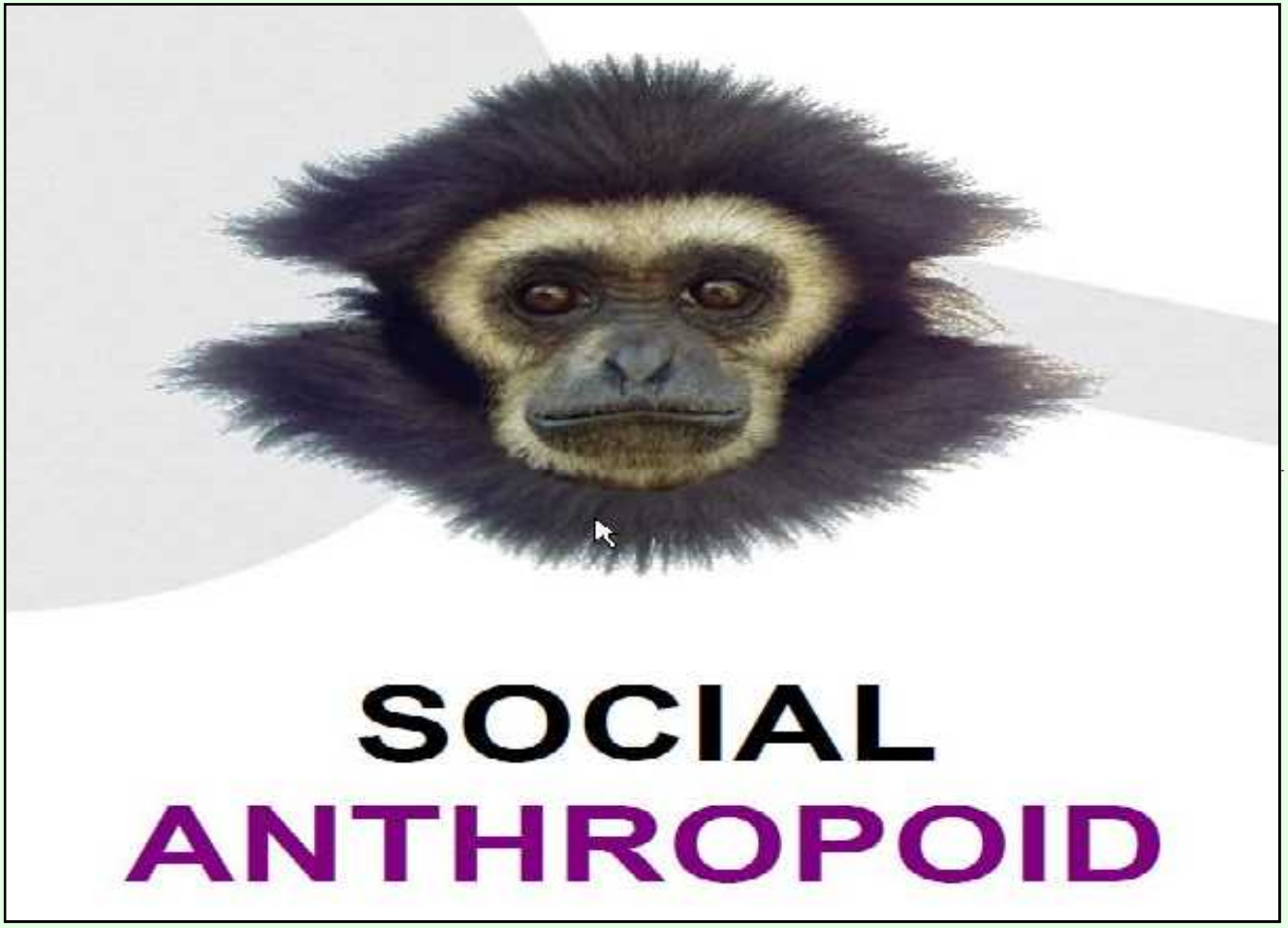

TDI listing
Browse by technology
Rows 15
Go
View Chart
Or, type in the TDI type you are interested in
TDI type Yahoo-Y-Cookie
Go
Technologies associated with Yahoo-Y-Cookie
Tech
Description
Yahoo
Yahoo provides various free web services including webmail, a web portal, a web directory, news, and mapping
1 - 1
Yahoo-Y-Cookie details
TDI Name Yahoo-Y-Cookie
Type TDI
Scope User
Description This is the username of the Yahoo! user who is logged in. The username is the first bit of the Yahoo! e-mail address (everything before the "@") if the user is on a primary domain - or the full email address if the user is on a secondary domain (see <a href=
[redacted]
>here</a> for more info on primary / secondary domains). In raw traffic, the username appears as part of the "Y=" cookie string, with "l=" before it and "/o" after it, obfuscated using a simple substitution code called ROT13. The QFDs automatically break out the relevant part of the cookie and turn it back into plain text for use as a TDI. If a Yahoo-Y-Cookie is seen with user agent Mozilla/4.0 (compatible; MSIE 5.5...) then that is probably the Yahoo! instant messaging client rather than a web browser.
Further info If the information above is blank, incomplete, or unhelpful, please email GTE Tech Tracking and demand answers.
BEGAL rules generating Yahoo-Y-Cookie events
Name
ID
EXP_Yahoo-Y-Cookie_0
11113
1 - 1
INFINITE MO
KARMA POLIC
MARBLED GE
MEMORY HOL
MUTANT BRO
SAMUEL PEPY
SOCIAL ANIMAL
SOCIAL ANTHROPOID user guide
TDI Database login using Corporate Directory password
Subscribe to the blog feed on the GCWeb dynamic home page.

events Samuel Pepys

- Pull through and upscaling of TR SPs.
  - Currently 43 bearers.
    - 14 from TR SP
    - 29 additional bearers from TPS (generating HTTP, TDI, Websearch, FTP and Squeal).
    - Circa 40 additional bearers just generating Squeal.
  - Approval to increase aperture to 100 bearers for all data-types.
  - Approval to increase user numbers to 200.

GCHQ

events



GCHQ

- What is Social Anthropoid?
  - SOCIAL ANTHROPOID is a converged comms database. It will allow you to see when your targets have communicated via phone, over the internet, or using converged channels (e.g., sending e-mails from a phone or making voice calls over the internet).

events

- What about the existing comms data-bases?
  - When SOCIAL ANTHROPOID contains all the necessary data and has all the core functionality of the legacy tools Social animal, HAUSTORIUM and SALAMANCA will be de-commissioned.

events

- What data is in Social Anthropoid??
  - All of Salamanca data (telephony)
  - Social animal data.
  - Instant Messenger.
  - Webmail. - SIP & H323 VOIP
  - Yahoo Voice
  - Blackberry
  - MMS
  - SMS (from Salamanca and other sources)
  - GTP (GPRS session set-ups)
  - And more..

GCHQ

- What about SMTP, POP3 and IMAP?
  - Starting to receive these data types now.
  - Capability deployed as part of HeartBeat 11.

events

Query input

Saved queries

You have 1 saved queries.

Test

- Queries will be automatically submitted to all instances of SOCIAL ANTHROPOID, SOCIAL ANIMAL and Convergec SOCIAL ANIMAL.
- For bulk queries, enter multiple selectors (one per line).
- If allow wildcards is ticked, * is treated as a multi-character wildcard (e.g. paul* will match paul123, paul456, paul*@yahoo.com will match paul123@yahoo.com but not paul123@hotmail.com). Unlike other QFDs, _ and \ have no special meaning (to query for a literal * sign, uncheck 'allow wildcards' rather than 'escaping' the wildcard.
- By default, results will be returned in which your input selector appears in either the User A or User B column (in SOCIAL ANIMAL terms is the 'actor' or the 'subject' with in the event). To return results in which your selector appears only as the active user, tick the 'Query active users only' checkbox
- Front-end processing normalises C2C selectors in various ways, including the removal of dots from the usernames of Gmail addresses. To get Gmail results, you will need to normalise your queries in the same way (e.g., search for badguy@gmail.com instead of bad.guy@gmail.com). Gmail itself ignores the dots so there is no danger of getting events for the wrong account. If in doubt, consult your local C2C technical.

Miranda 2014C

C2C Priority & Purpose 1kS

HRA Justification testing

Search period (optional): to

Filter results by matched selectors prior to display

Allow wildcards

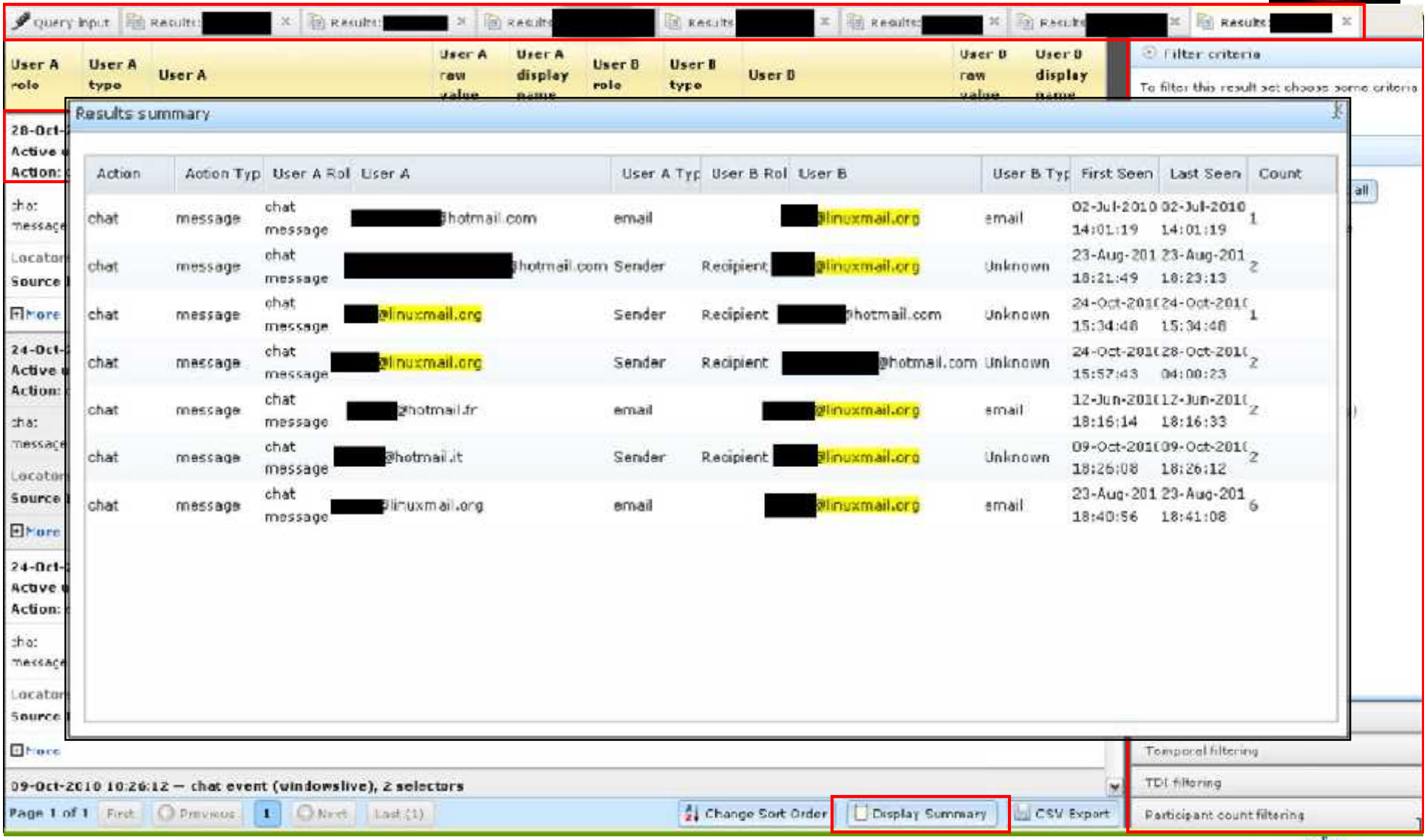
Query active users only

Save Query | Submit query

SECRET STRAP1

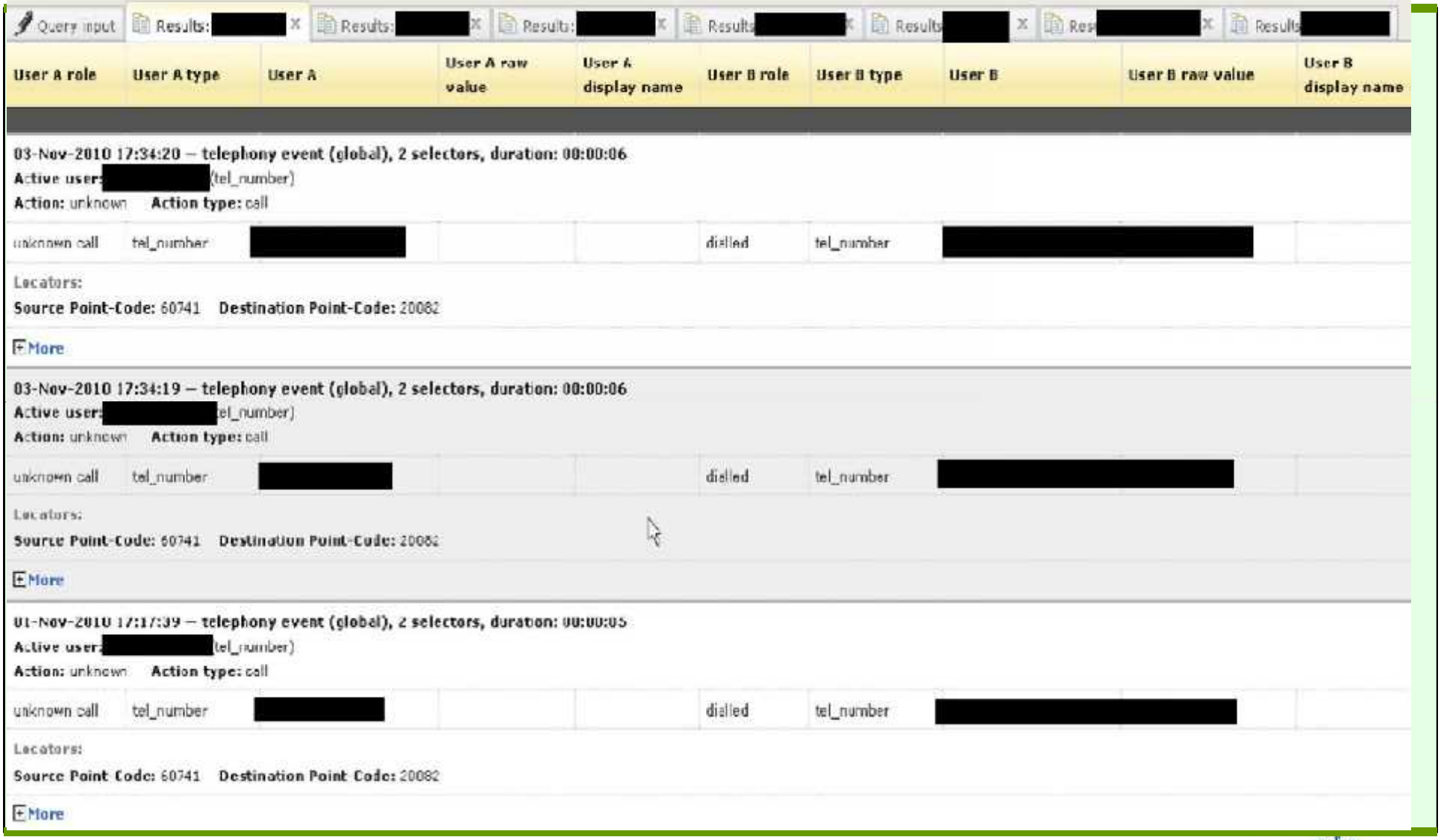
GCHQ

events C2C



SECRET STRAP1

Telephony in Santhropoid



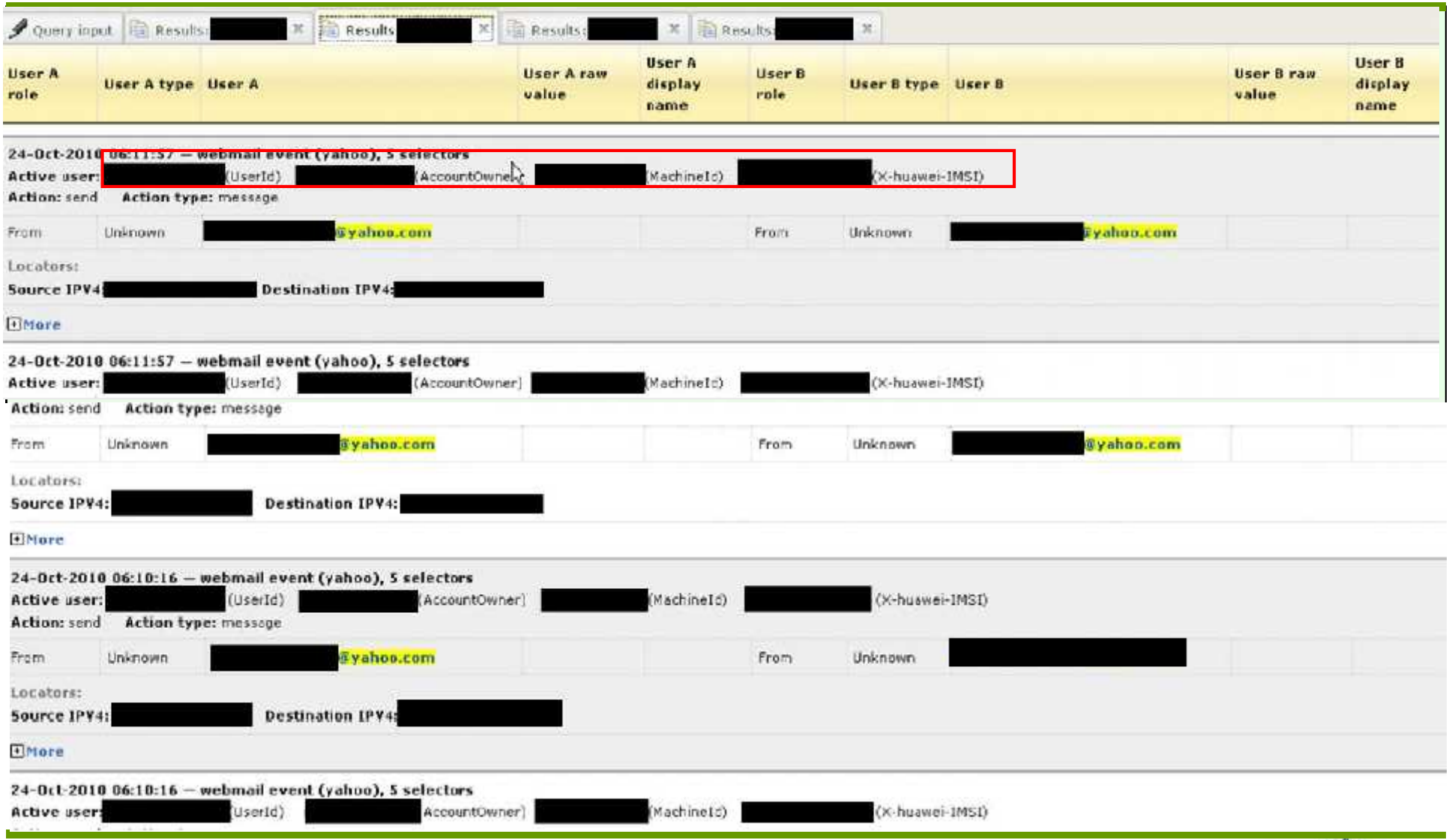
events Convergence - GTP tunnel

Query input | Results: [redacted] | Results [redacted] | Results: [redacted]

User A role	User A type	User A	User A raw value	User A display name	User B role	User B type	User B	User B raw value	User B display name
01-Nov-2010 07:16:06 – tunnel event (gtp), 3 selectors									
Active user: [redacted] (imsi) [redacted] (tel_number) [redacted] (blackberry_mid_pin)									
Action: create Action type: tunnel									
create tunnel	imsi	[redacted]				not_available	not_available		
create tunnel	tel_number	[redacted]				not_available	not_available		
create tunnel	blackberry_mid_pin	[redacted]				not_available	not_available		
Locators:									
Source IPV4: [redacted] Source SgsnAddress: [redacted] Destination IPV4: [redacted]									
GTP Convergence Data:									
This event represents the creation of GTP tunnel [redacted]									
View all events from this GTP tunnel									
More									
01-Nov-2010 04:28:18 – tunnel event (gtp), 3 selectors									
Active user: [redacted] (imsi) [redacted] (tel_number) [redacted] (blackberry_mid_pin)									
Action: create Action type: tunnel									
create tunnel	imsi	[redacted]				not_available	not_available		
create tunnel	tel_number					not_available	not_available		
create tunnel	blackberry_mid_pin					not_available	not_available		
Locators:									
Source IPV4: [redacted] Source SgsnAddress [redacted] Destination IPV4: [redacted]									
GTP Convergence Data:									
This event represents the creation of GTP tunnel [redacted]									
View all events from this GTP tunnel									

GCHQ

Convergence – Leaky Gateways



events OSN

Query input | Results [redacted] | Results: [redacted] | Results [redacted] | Results [redacted] | Results: [redacted] | Results [redacted] | Resu [redacted]

User A role	User A type	User A	User A raw value	User A display name	User B role	User B type	User B	User B raw value	User B display name

15-Jul-2010 21:41:42 — SOCIAL ANIMAL event (Facebook), 2 selectors
Active user: [redacted] (Facebook-uid)
Action: chat **Action type:** message

chat message	Facebook-uid	[redacted]				Facebook-uid	[redacted]		

Locators:
Source IPV4: [redacted]

More

15-Jul-2010 21:41:42 — SOCIAL ANIMAL event (Facebook), 2 selectors
Active user: [redacted] (Facebook-uid)
Action: alias **Action type:** user

alias user	Facebook-uid	[redacted]				email	[redacted]		

Locators:

More

15-Jul-2010 21:41:37 — SOCIAL ANIMAL event (Facebook), 2 selectors
Active user: [redacted] (Facebook-uid)
Action: chat **Action type:** message

chat message	Facebook-uid	[redacted]				Facebook-uid	[redacted]		

Locators:
Source IPV4: [redacted]

More

15-Jul-2010 21:41:33 — SOCIAL ANIMAL event (Facebook), 2 selectors
Active user: [redacted] (Facebook-uid)
Action: chat **Action type:** message

GCHQ

events

- Looks good, When can I have an account?
  - Santhropoid is currently in the second stage of UAT.
  - We currently have 200 users representing all areas of the business.
  - Aiming to be in a position to release Santhropoid to the masses in early January.

events New data sources

- LUSTRE – new data-source available in MB. Good for North Africa.

- Source field – This will enable new non-routine data-sources to be added to the QFD's.
  - CNE
  - JTRIG – GLASSBACK data used for test case.
  - COLLATERAL

GCHQ

New Loaders

- New loaders deployed to MB and HR Map, improvements to KP.
  - Latency of the data in the QFDs has been greatly reduced, now around 12 hours.
    - Each instance of MB can now ingest 8 billion events per-day (total 32 billion)
  - Some QFDs were previously 1-5 days behind.
  - Query performance during loading has also been improved.

events

GUIDING LIGHT QFD

Presented by [redacted]
(Guiding Light SU)

GCHQ

events

What is GUIDING LIGHT?

- New QFD developed in August 2010 by TDB-Events.
- Primary objective:

> "To understand the traffic seen on the Next Gen Events bearers."

GCHQ

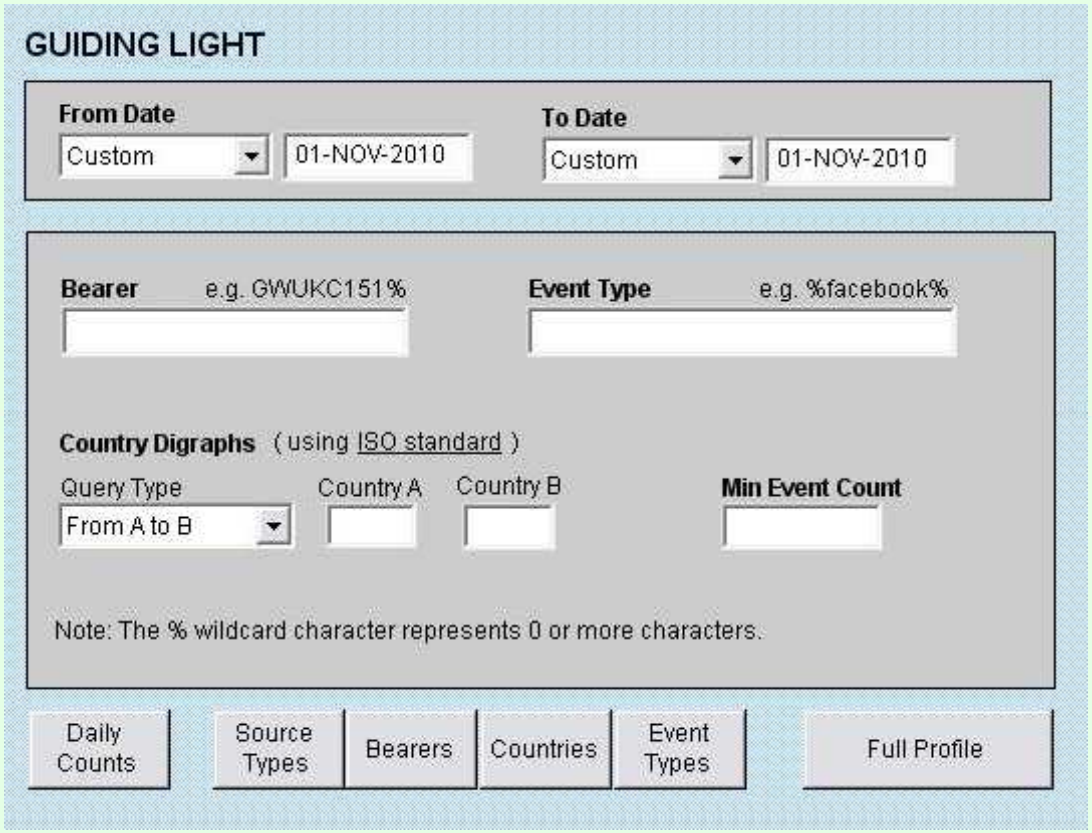
events

What can it do for me?

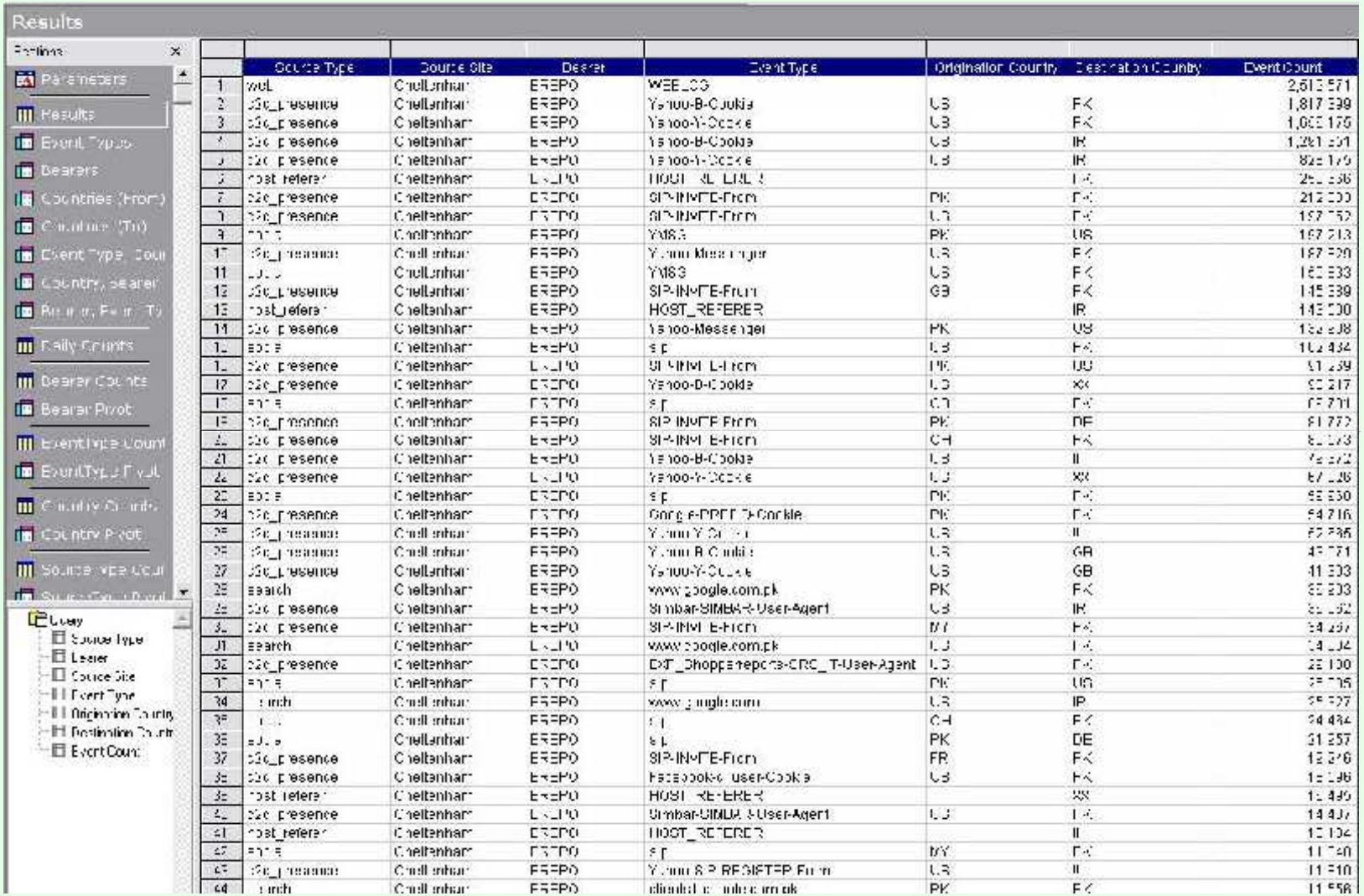
General Questions:

- Given a case notation, what are the TDI types that are found on it?
- Given a TDI type/subset, which bearers produce the highest number of events?
- What type of traffic is on which bearers and where is it coming from?
- Which bearers provide the most amount of traffic type *x* from place *y*?

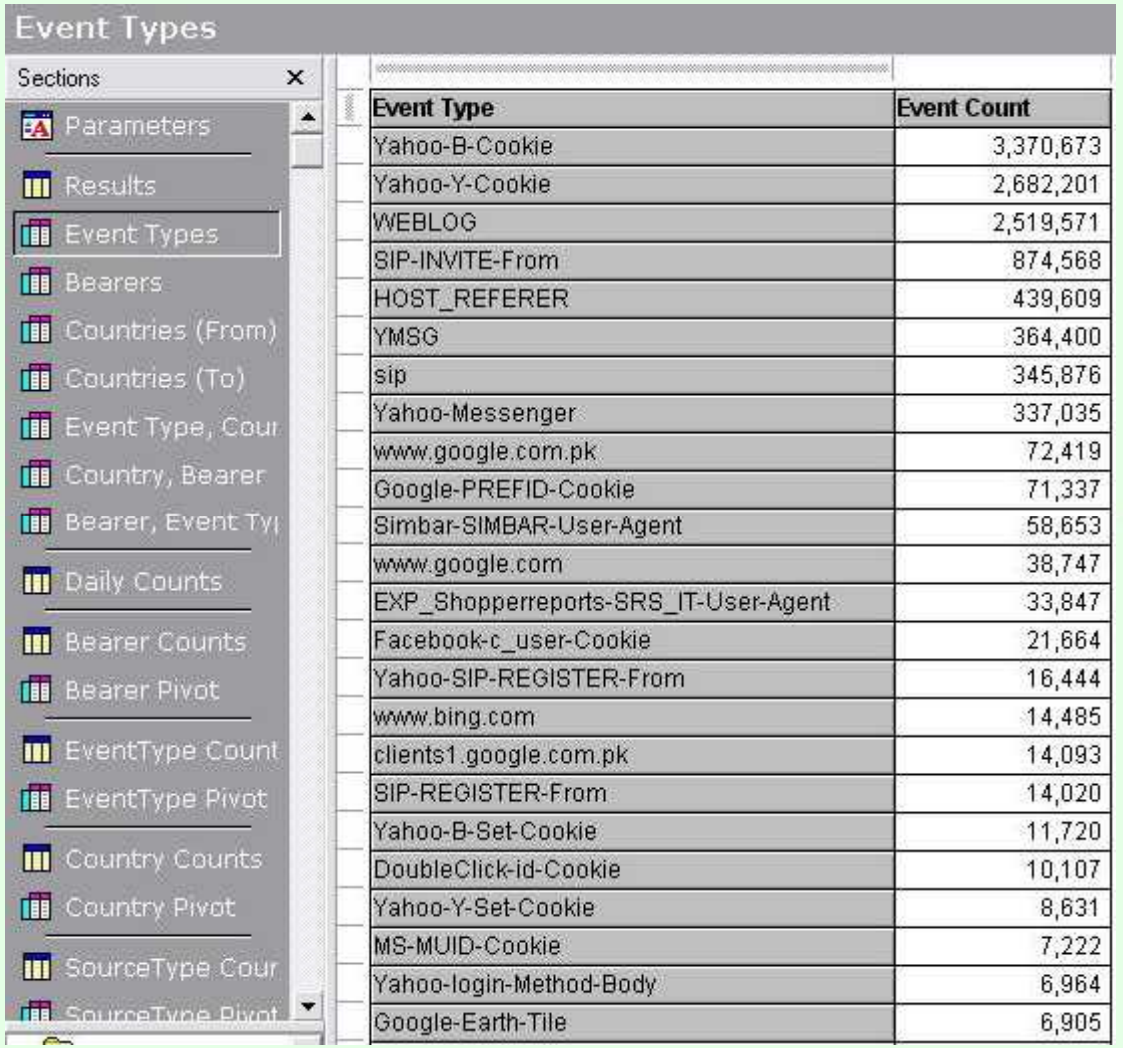
Front End Interface



Results - Full Profile Query

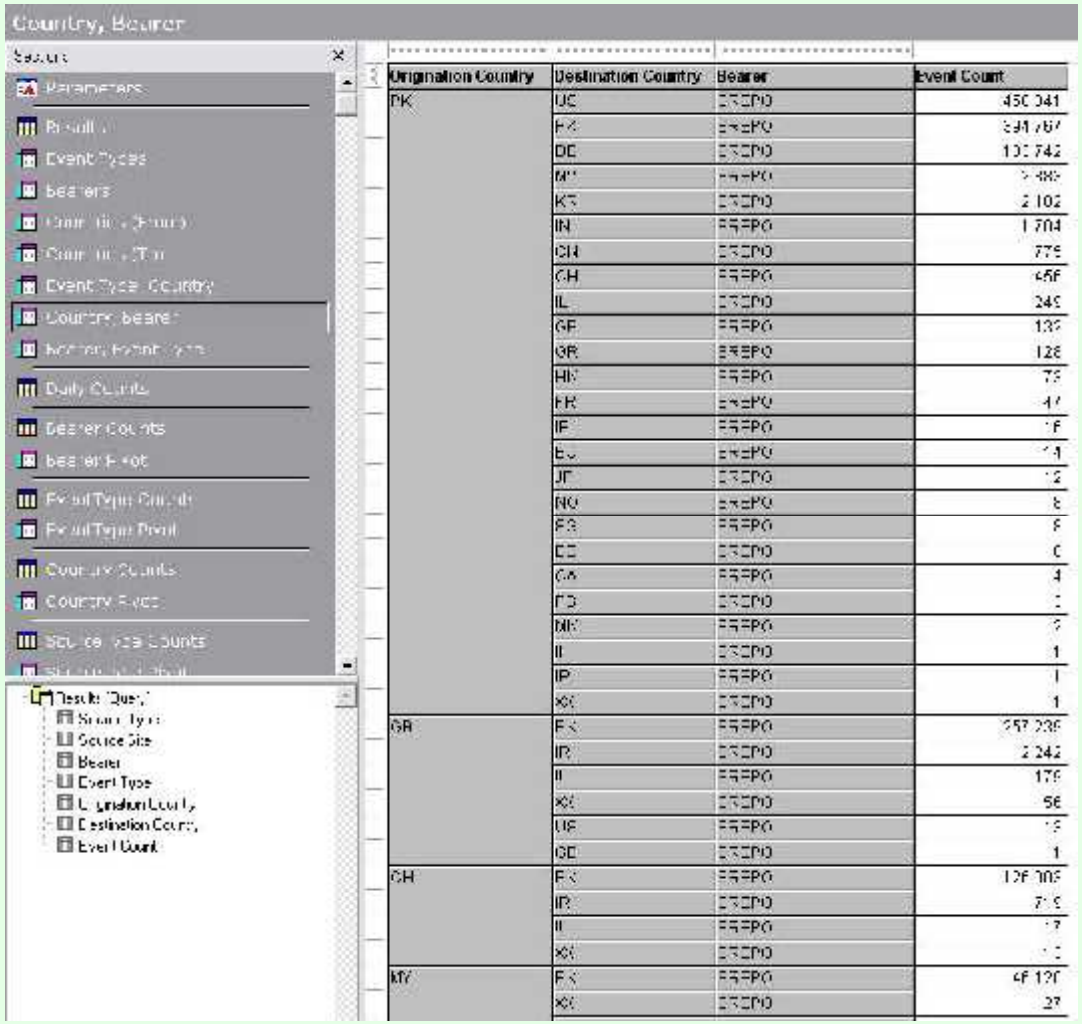


Results Pivot: Event Types



Event Type	Event Count
Yahoo-B-Cookie	3,370,673
Yahoo-Y-Cookie	2,682,201
WEBLOG	2,519,571
SIP-INVITE-From	874,568
HOST_REFERER	439,609
YMSG	364,400
sip	345,876
Yahoo-Messenger	337,035
www.google.com.pk	72,419
Google-PREFID-Cookie	71,337
Simbar-SIMBAR-User-Agent	58,653
www.google.com	38,747
EXP_Shopperreports-SRS_IT-User-Agent	33,847
Facebook-c_user-Cookie	21,664
Yahoo-SIP-REGISTER-From	16,444
www.bing.com	14,485
clients1.google.com.pk	14,093
SIP-REGISTER-From	14,020
Yahoo-B-Set-Cookie	11,720
DoubleClick-id-Cookie	10,107
Yahoo-Y-Set-Cookie	8,631
MS-MUID-Cookie	7,222
Yahoo-login-Method-Body	6,964
Google-Earth-Tile	6,905

Results Pivot: Countries (From)



Origination Country	Destination Country	Bearer	Event Count
PK	US	[illegible]	450 341
	FR	[illegible]	[illegible]
	DE	[illegible]	100 742
	[illegible]	[illegible]	[illegible]
	KR	[illegible]	2 102
	IN	[illegible]	1 704
	CN	[illegible]	779
	CH	[illegible]	456
	IL	[illegible]	249
	GB	[illegible]	139
	GR	[illegible]	128
	HK	[illegible]	79
	ER	[illegible]	47
	IE	[illegible]	18
	BE	[illegible]	14
	JP	[illegible]	12
	NO	[illegible]	8
	ES	[illegible]	6
	CZ	[illegible]	6
	CA	[illegible]	4
	FI	[illegible]	2
	MX	[illegible]	2
	IL	[illegible]	1
	IP	[illegible]	1
	XX	[illegible]	1
GR	ES	[illegible]	257 236
	IR	[illegible]	2 242
	IL	[illegible]	179
	XX	[illegible]	56
	US	[illegible]	19
	GB	[illegible]	1
CH	ES	[illegible]	126 305
	IR	[illegible]	7 [illegible] 9
	IL	[illegible]	17
	XX	[illegible]	1 2
MY	ES	[illegible]	46 126
	XX	[illegible]	27

Recent Enhancements

- Data from Bude (RPC)
  - Including data from SWORDPLAY
- New fields
  - PDDG
  - SIGAD
  - SSDG

Future Enhancements

Near future:

- Adding BROAD OAK Targeting data
- Incorporating MI functionality from REFORMER (where appropriate!)
- Adding more feeds. (Ongoing)

Longer term:

- Adding Cipher and eAD MI information
- Linkage into ARTEMIS (or its successor)

Any Questions

?