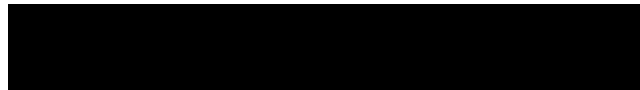
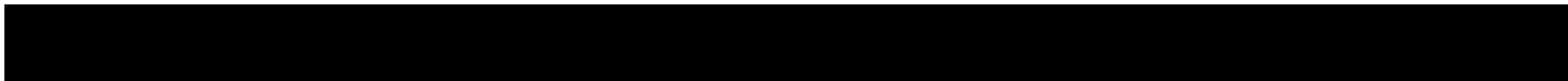




SSO Collection Optimization

Core SSO Team:





Address Books

- Email address books for most major webmail are collected as stand-alone sessions (no content present*)
- Address books are repetitive, large, and metadata-rich
- Data is stored multiple times (MARINA/MAINWAY, PINWALE, CLOUDs)
- Fewer and fewer address books attributable to users, targets
- Address books account for ~ 22% of SSO's major accesses (up from ~ 12% in August)

Access (10 Jan 12)	Total Sessions	Address Books	Provider	Collected	Attributed	Attributed%
US-3171	1488453	237067 (16% of traffic)	Yahoo	444743	11009	2.48%
DS-200B	938378	311113 (33% of traffic)	Hotmail	105068	1115	1.06%
US-3261	94132	2477 (3% of traffic)	Gmail	33697	2350	6.97%
US-3145	177663	29336 (16% of traffic)	Facebook	82857	79437	95.87%
US-3180	269794	40409 (15% of traffic)	Other	22881	1175	5.14%
US-3180 (16 Dec 11)	289318	91964 (32% of traffic)	TOTAL	689246	95086	13.80%
TOTAL	3257738	712366 (22% of traffic)				

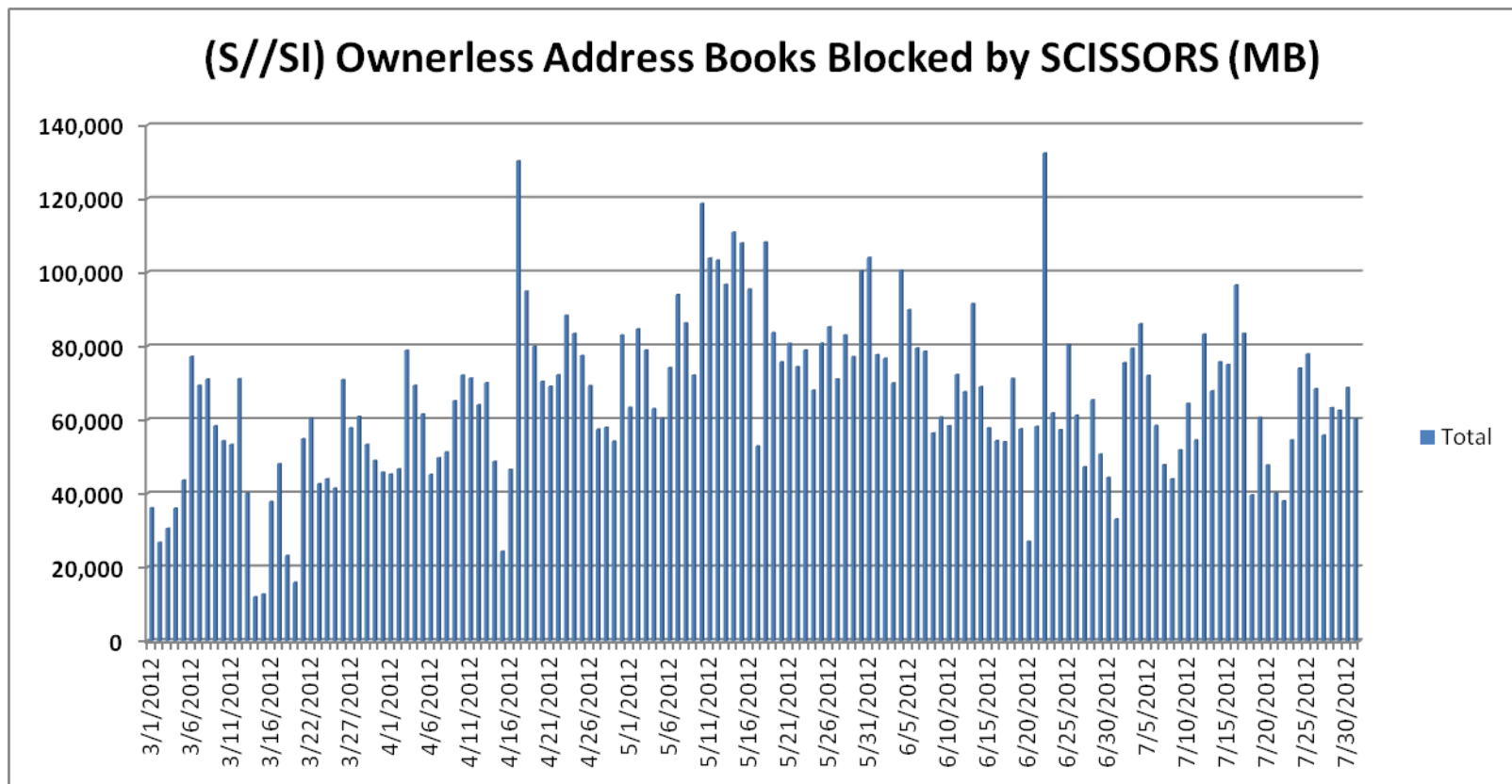


Address Books

- Enabled in SCISSORS for various SSO sites:
 - **JPMQ** (metadata: QMPJ) - DS-200B (MUSCULAR) 29 Feb 2012
 - **DGOT** (metadata: TOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - **DGOD** (metadata: DOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - **SPNN** (metadata: NNPS) - US-3180 (SPINNERET) 03 May 2012
 - **EGLP** (metadata: PLGE) - US-3145 (MOONLIGHTPATH) 08 May 2012

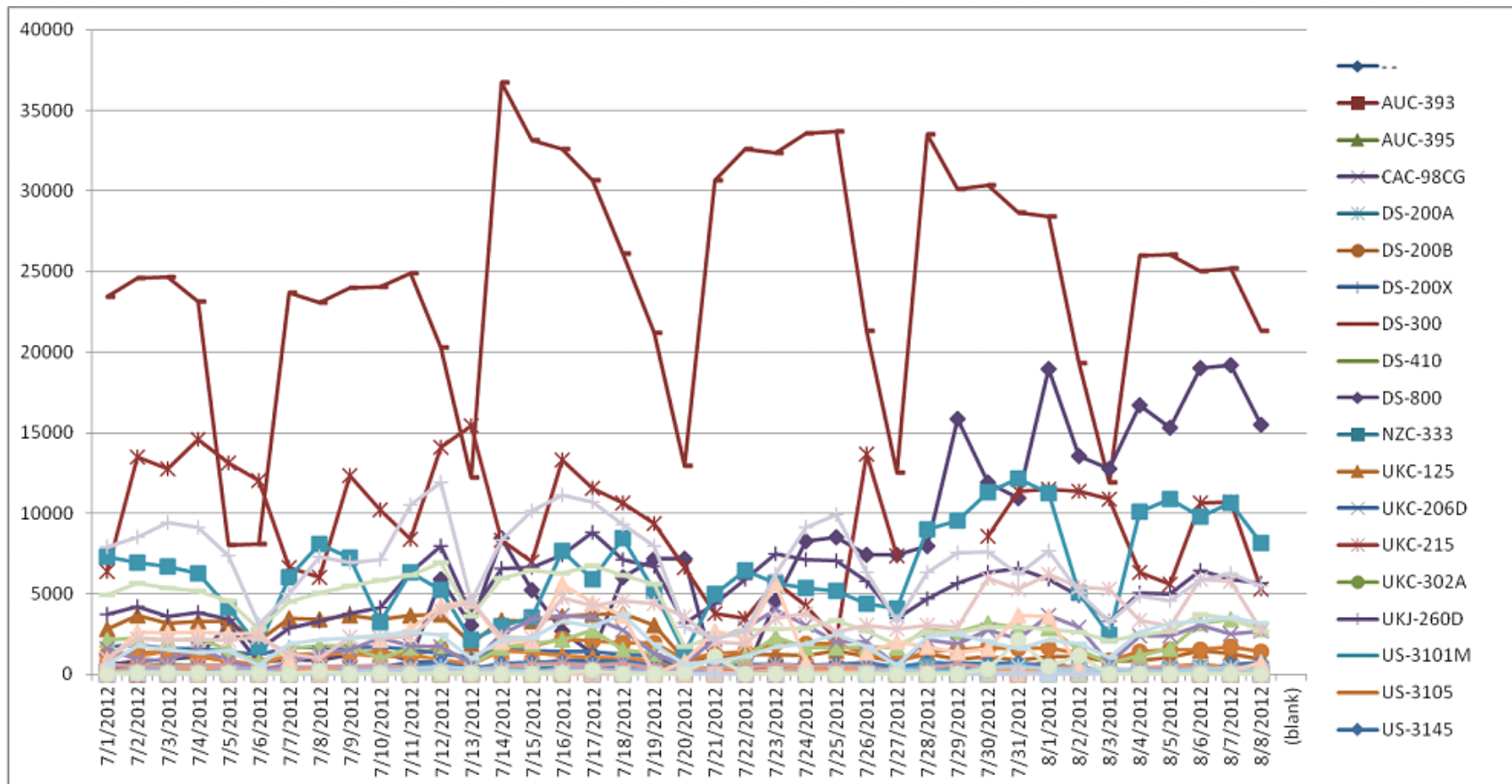


Address Books



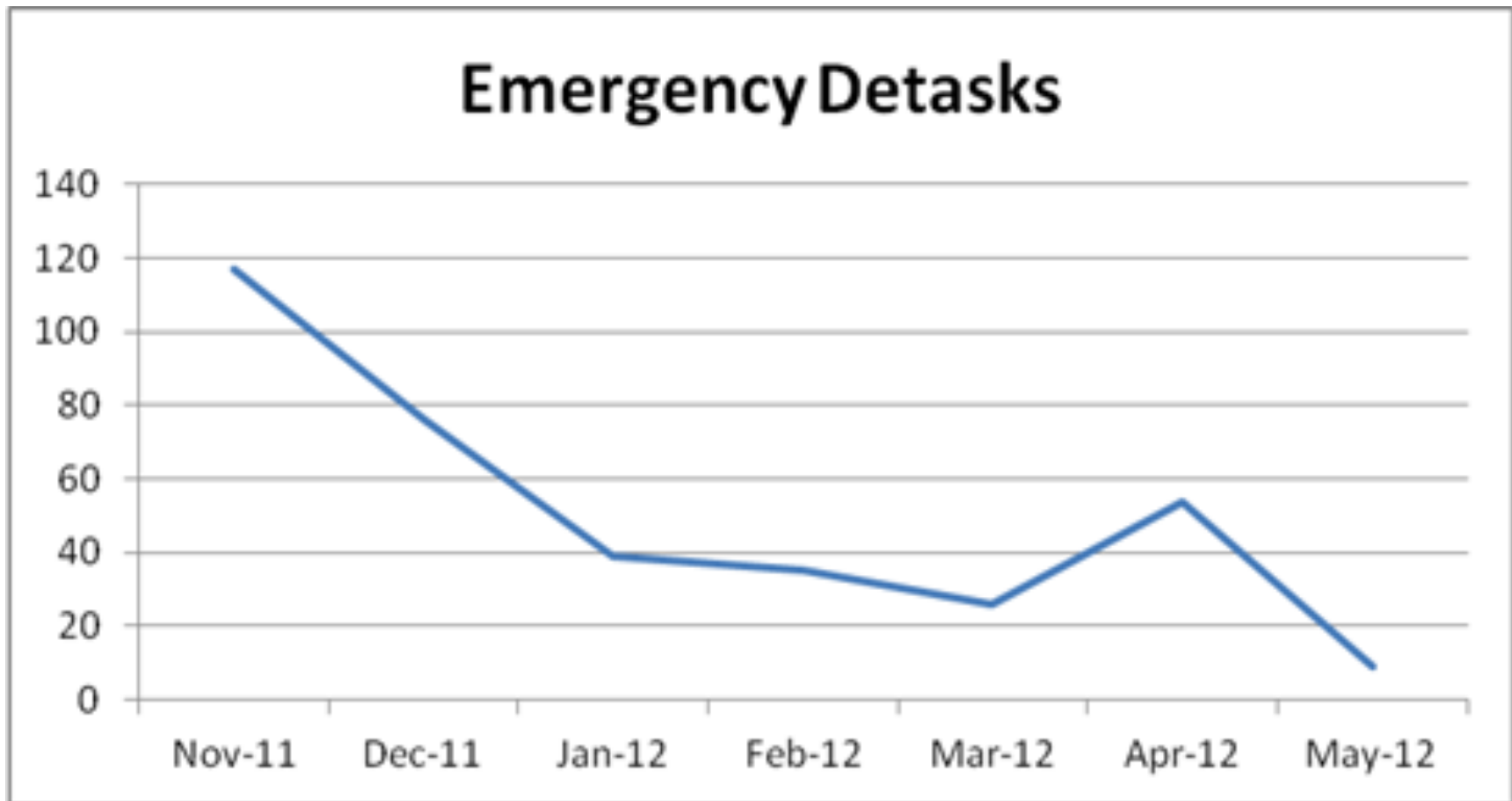


Address Books





Selector Detasks





So What?

- Store less of the wrong data
 - 20% reduction (so far) in content to long-term repositories
 - Data still resides at site for SIGDEV
- Increase data variety
 - Hole left by “wrong data” filled with more “right data”
 - More signals and case notations can be tasked at site
- Shifting collection philosophy at NSA
 - “Memorialize what you need” versus “Order one of everything off the menu and eat what you want”

WIKI: https://wiki.nsa.ic.gov/wiki/Collection_Optimization

XKEYSCORE: fingerprint/defeats/atrouter and fingerprint/defeats/atxks

Collection Optimization

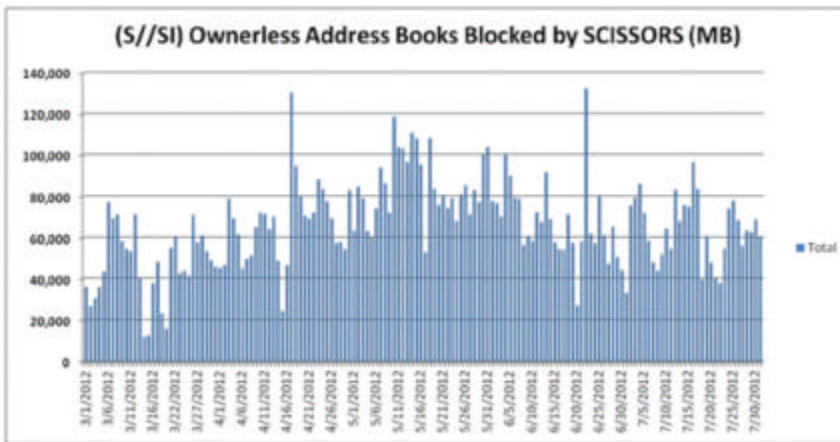
7 Pages - Contributed by Matt DeLong, Washington Post - Oct 14, 2013

This is another presentation on problems with NSA overcollection.

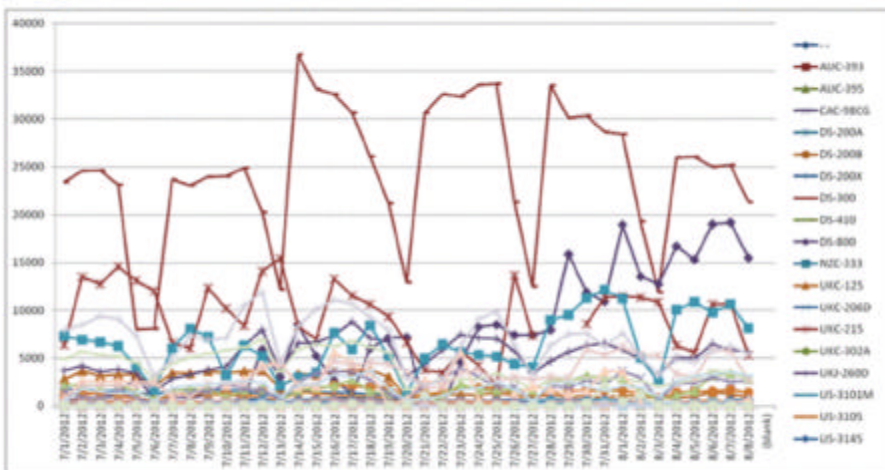
SCISSORS (p. 3)

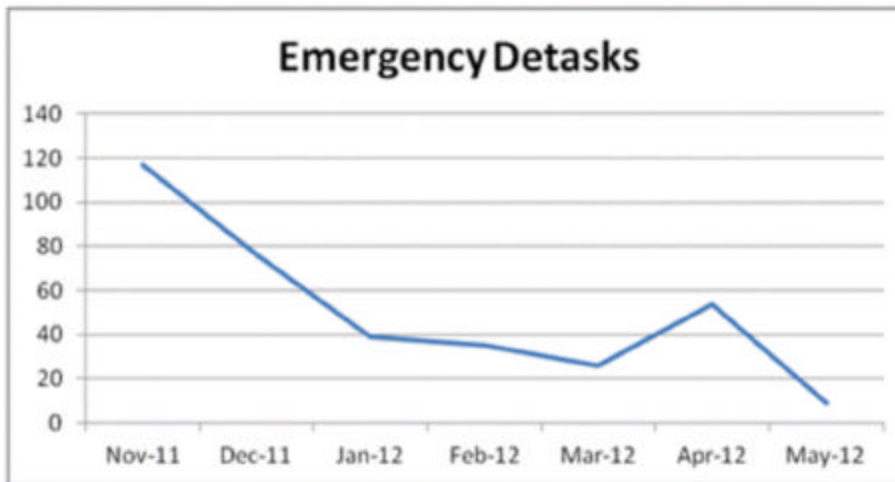
- Enabled in SCISSORS for various SSO sites:
 - JPMQ (metadata: QMPJ) - DS-200B (MUSCULAR) 29 Feb 2012
 - DGOT (metadata: TOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - DGOD (metadata: DOGD) - US-3171 (DANCINGOASIS) 13 Mar 2012
 - SPNN (metadata: NNPS) - US-3180 (SPINNERET) 03 May 2012
 - EGLP (metadata: PLGE) - US-3145 (MOONLIGHTPATH) 08 May 2012

Ownerless address books blocked by SCISSORS (p. 4)



Ownerless address books blocked, by points of access (p. 5)



Emergency detasks (p. 6)**SIGDEV** (p. 7)

- Store less of the wrong data
 - 20% reduction (so far) in content to long-term repositories
 - Data still resides at site for SIGDEV

"Shifting collection philosophy at NSA" (p. 7)

- Shifting collection philosophy at NSA
 - "Memorialize what you need" versus "Order one of everything off the menu and eat what you want"