
The CRFometer

A data-flow monitoring system for providing DataFax users and administrators detailed information about incoming CRF traffic

Craig Agamemnon Magaret
Statistical Center for HIV/AIDS Research and Prevention
Fred Hutchinson Cancer Research Center
cmagaret@scharp.org

SCHARP's DataFax Operations

- 12+ active clinical trials and observational studies
- Receives between 3,000-18,000 faxed CRFs weekly
- CRFs are submitted from 50+ sites across five continents
- Keeping everything running smoothly is a logistical challenge

System needed for tracking CRFs

- Identify problems when they occur: (e.g., “Is Site X having technical difficulties?” “Is our DataFax server tweaking out again?”)
- Measure trends of fax volume over time (“What workload do we need to prepare for?” “What is the best time/day to take down the server?”)
- Provide information in real-ish time

Part 1: The Boring Part

- Every system needs an under-appreciated “back-end” that does all the work and receives no glory
- Cron job runs hourly; checks fax_log for new entries since last execution
- Information parsed from fax_log entries:
 - Sender ID
 - Number of pages transmitted
- Recorded in flat-file, pipe-delimited database

The Configuration File

- “Which sender IDs correspond to which columns?” Cron job sorts them out via configuration file
- Config file: one row per sender ID
- The cron job recognized new sender IDs and appends them to config file
- Config file is designed to include optional information about sites (e.g., name of clinic, et al), which is added manually

The Configuration File: An Example

```

206 364 9514|325|Kindred Hospital|Seattle, USA|phone
314 298 1504|326|Unknown (314 298 1504)||phone
314-771-3816|327|Unknown (314-771-3816)||phone
df-lun-kamwala@cidrz.org|328|CIDRZ-Kamwala|Lusaka, Zambia|email|HPTN
df-lun-chilenje@cidrz.org|329|CIDRZ-Chilenje|Lusaka, Zambia|email|HPTN
cidrz-kamwala@cidrz.org|330|CIDRZ-Kamwala|Lusaka, Zambia|email|HPTN
cidrz-matero@cidrz.org|331|CIDRZ-Matero|Lusaka, Zambia|email|HPTN
cidrz-george@cidrz.org|332|CIDRZ-George|Lusaka, Zambia|email|HPTN
2685196|333|Unknown (2685196)||phone
6172774288|334|Unknown (6172774288)||phone
hptnfx@mrca.ac.za|335|MRC|Hlabisa, South Africa|email|HPTN
support@scharp.org|336|SCHARP IT Support||email
031|337|Unknown (031)||phone
jayakumar@yrgcare.org|338|YRG Care|Chennai, India|email
206 628 8576|339|Unknown (206 628 8576)||phone
914 664 7650|340|Unknown (914 664 7650)||phone
MUJHUFAX@MUJHU.ORG|341|JHU/Mulago|Kampala, Uganda|email|HPTN
DF_LIM2@impactaperu.org|342|Impacta-Ponce|Lima, Peru|email
DF_PUC1@impactaperu.org|343|Impacta-Pucallpa|Pucallpa, Peru|email
DF_HRE1@uz-ucsf.co.zw|344|Phillips (F104L, ground floor)|Harare, Zimbabwe|email|
hptn_fax2@uz-ucsf.co.zw|345|Phillips (Unknown Fax Machine)|Harare, Zimbabwe|email|
df_hre2@uz-ucsf.co.zw|346|Phillips (Unknown Fax Machine)|Harare, Zimbabwe|email|
DF_HRE2@uz-ucsf.co.zw|347|Phillips (F104L, IT room)|Harare, Zimbabwe|email|HPTN
hvtu-dr@codetel.net.do|348|Instituto Dermatologico|Santo Domingo, Dominican Repu
3902671|349|Unknown (3902671)||phone
faxMP1195@scharp.org|350|SCHARP (ME-11 Reception)|Seattle, USA|email
9899762|351|Unknown (9899762)||phone
2125799849|352|Unknown (2125799849)||phone
250 503233|353|Unknown (250 503233)||phone
kmoyo@bhp.org.bw|354|Unknown (kmoyo@bhp.org.bw)||

```

What Does This Give Us?

- The cron job creates a flat-file database of incoming CRF traffic, tracking “how many faxes were sent by whom, when”
- Database can be imported into Excel and other data-processors
- The cron job can be executed at any frequency to meet desired resolution
- Obvious problem: database only accumulates from time you start cron job; what to do if you want your database to predate the cron job?

The Harvester Tool

- If fax info is present in the fax_log, it can be “harvested” into a working database
- Since date entries in fax_log do not represent local server time, time/date errors may occur
- Harvester creates its own database and config file; these need to be merged with those created by cron job (a task that currently needs to be performed manually)

Caveats

- Since database is a collection of time-series data, missing datapoints can cause problems: I.e., If your cron job doesn't run, bad things can happen
- Merging the harvested database with the ongoing database can seem like trying to cram an angry cat into a lunchbox
- Tools have been created to address these problems

Part 2: The Nifty Part

- “Gee, Craig, this is nice and everything, but so far I'd rather be eating wet newspapers than listening to you prattle on about this system stuff for – oh God, is it really another eleven minutes? Where's a tranquilizer gun when you need one?”
- Okay, smarty-pants: With the data, you can impress your colleagues with cool charts and graphs

Adventures In Data Visualization

- As mentioned earlier, database can be imported into Excel/Gnuplot/whatever, and tortured to your heart's content
- For cool graphs in real time, the web seems to be an ideal platform
- Graphs/tables are created/updated with every database update; included in web pages via SSI
- Display CRF volume over time, activity from individual sites, et al.

SCHARP's CRFometer Web Page



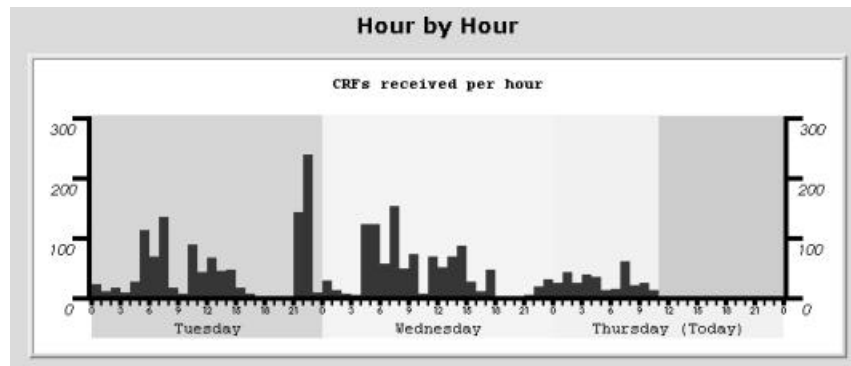
Cool Graphs

SCHARP DataFax at a Glance

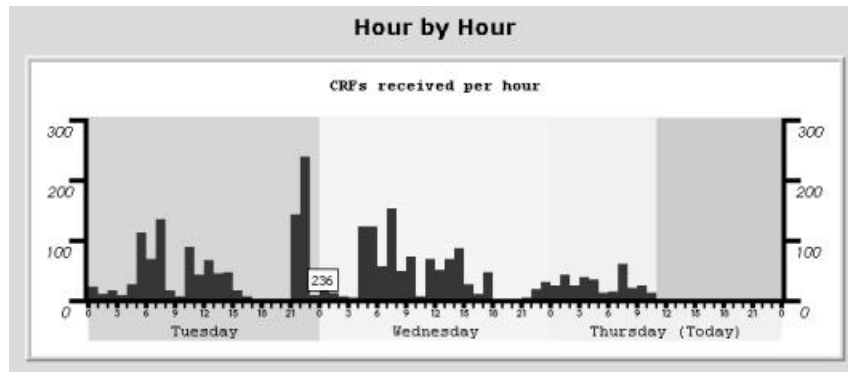
In the last	We have received this many pages
Hour	10
24 Hours	655
7 Days	4301
30 Days	23151
Year	404438
Eternity	3451674

Cool Graphs

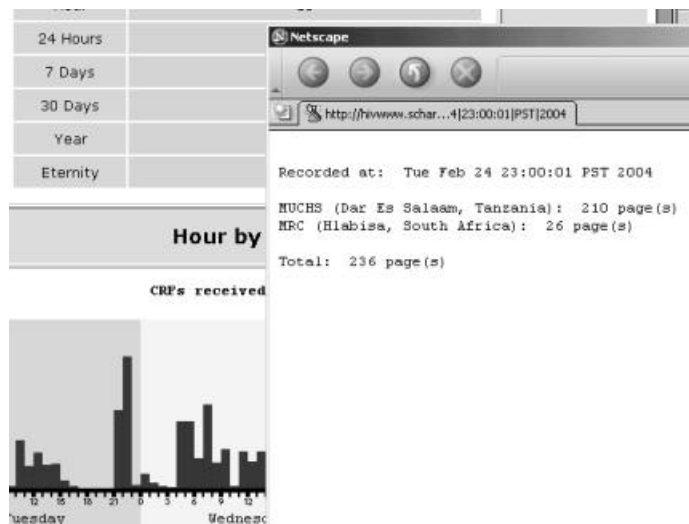
Hour by Hour



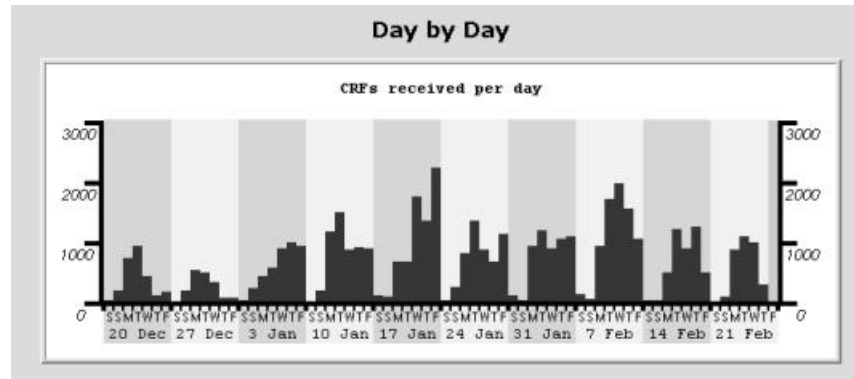
Cool Graphs - Data on Display



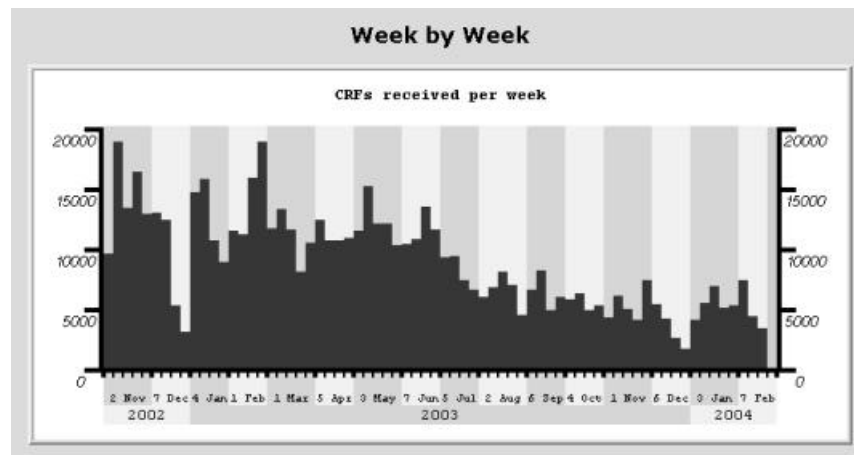
Cool Graphs - Getting Specifics



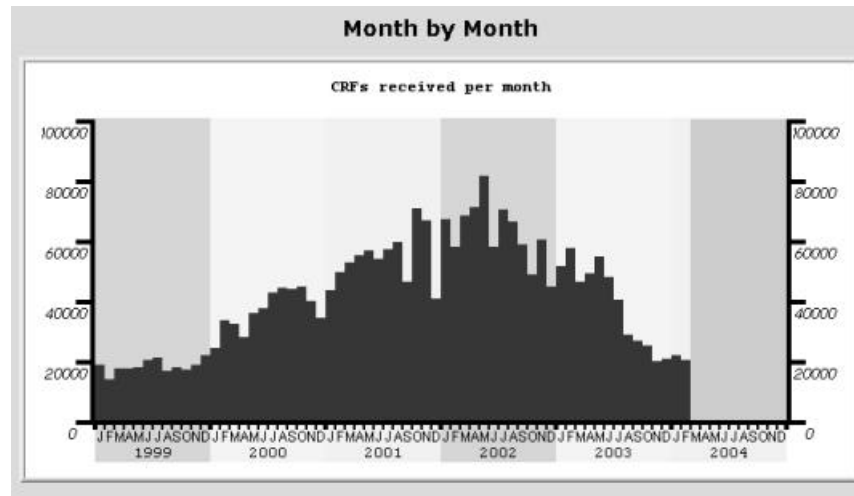
Cool Graphs



Cool Graphs



Cool Graphs



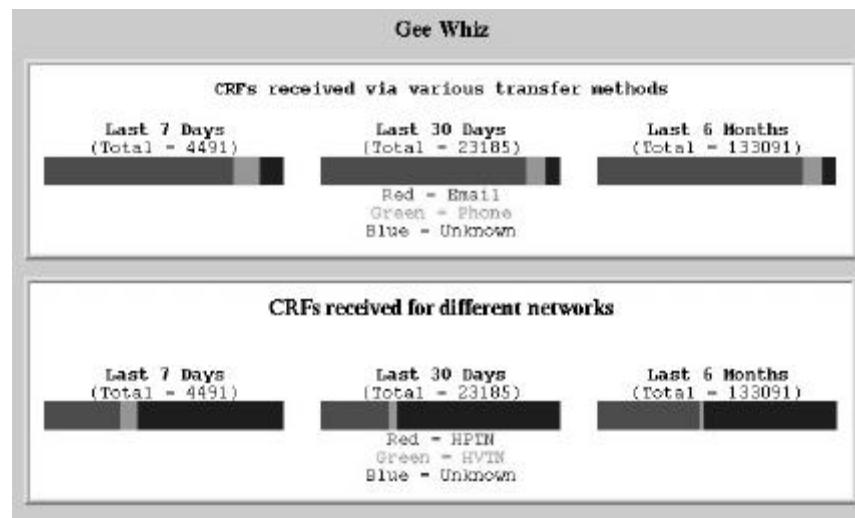
Cool Graphs

Faxes by Site

CRFs recently received from sites, and the hiatus since our last reception

Site	Hiatus	20 Feb	21 Feb	22 Feb	23 Feb	24 Feb	25 Feb	26 Feb
SCHARP (ME-11 Reception) (Seattle, USA)	0h	-	-	-	14	7	31	2
Unknown (6172774288)	0h	3	-	-	5	-	-	12
DataBuddy Test Unit (Seattle, USA)	0h	-	-	-	28	40	27	22
Impacta (Lima, Peru)	1h	36	-	-	37	11	36	7
Fenway Community Health (Boston, USA)	3h	2	-	-	-	4	46	3
MRC (Durban, South Africa)	6h	53	-	-	27	70	101	76
HIVNET Project (Lilongwe, Malawi)	11h	75	-	-	58	6	19	-
Impacta-Pucallpa (Pucallpa, Peru)	17h	-	-	-	41	-	48	-
MUCHS (Dar Es Salaam, Tanzania)	1d	-	-	-	-	349	-	-
PHRU (Johannesburg, South Africa)	2d	4	-	-	-	4	-	-
JCRC (Kampala, Uganda)	5d	85	-	-	-	-	-	-
APPEAL Study (Seattle, USA)	7d	-	-	-	-	-	-	-
CIDRZ-Kamwala (Lusaka, Zambia)	13d	-	-	-	-	-	-	-
Unknown (2125799849)	20d	-	-	-	-	-	-	-
Unknown (3902671)	28d	-	-	-	-	-	-	-

Gee-Whiz Graphs™



Future Plans

- Additional minor tools for back-end
- Graphs of “typical traffic”, with arithmetic means of hourly and daily volumes
- Ability to drill down into all graphs
- Custom Graph Generator, using user-supplied parameters via web interface
- Additional Gee-Whiz graphs
- Detailed site histories, w/rampant hyperlinking
- Programmable alarms
- Make sandwiches

Obligatory Mentioning of Contributors So They Won't Hate Me

- Web-design assistance: Shane Coultas
- Hand-Waving and Cabalistic DataFax Sorcery: Darryl Pahl
- Feature and Design Feedback: Roy Obenchain, Rana Bonnice, Melita Romasco, Tom Frasier, Deborah Donnell, Len Quiat, Alice Fisher