# URL Block List 2.0 - Expanding Data Sets for Precision Reporting and Deeper Data Mining

Patrick Cain

pcain@apwg.org

pcain@coopercain.com


APWG

Unifying the
Global Response
to Cybercrime

# Overview

- The APWG has been accepting phishing URLs for a long time
  - Expanded to accept ‘domains’ a few years ago
- Metrics and reports are generated from this data
- But ‘URL’ doesn’t give us:
  - DNS Data
  - IP Address
  - Actual lure or message
- Can we expand the UBL to include more data?
  - Email addresses, fone #s, ?, ?

# The Processing challenge

- We accept data via email, or email, or email.
  - This may not work with an expanded data set
- Divided submittors into low-volume (LV) and high-volume (HV)
  - Devise a web page for the LV
  - Develop scripts for the HV
- You get the data via a ‘pull’ from a repository
- Hard to search – manually or scriptily

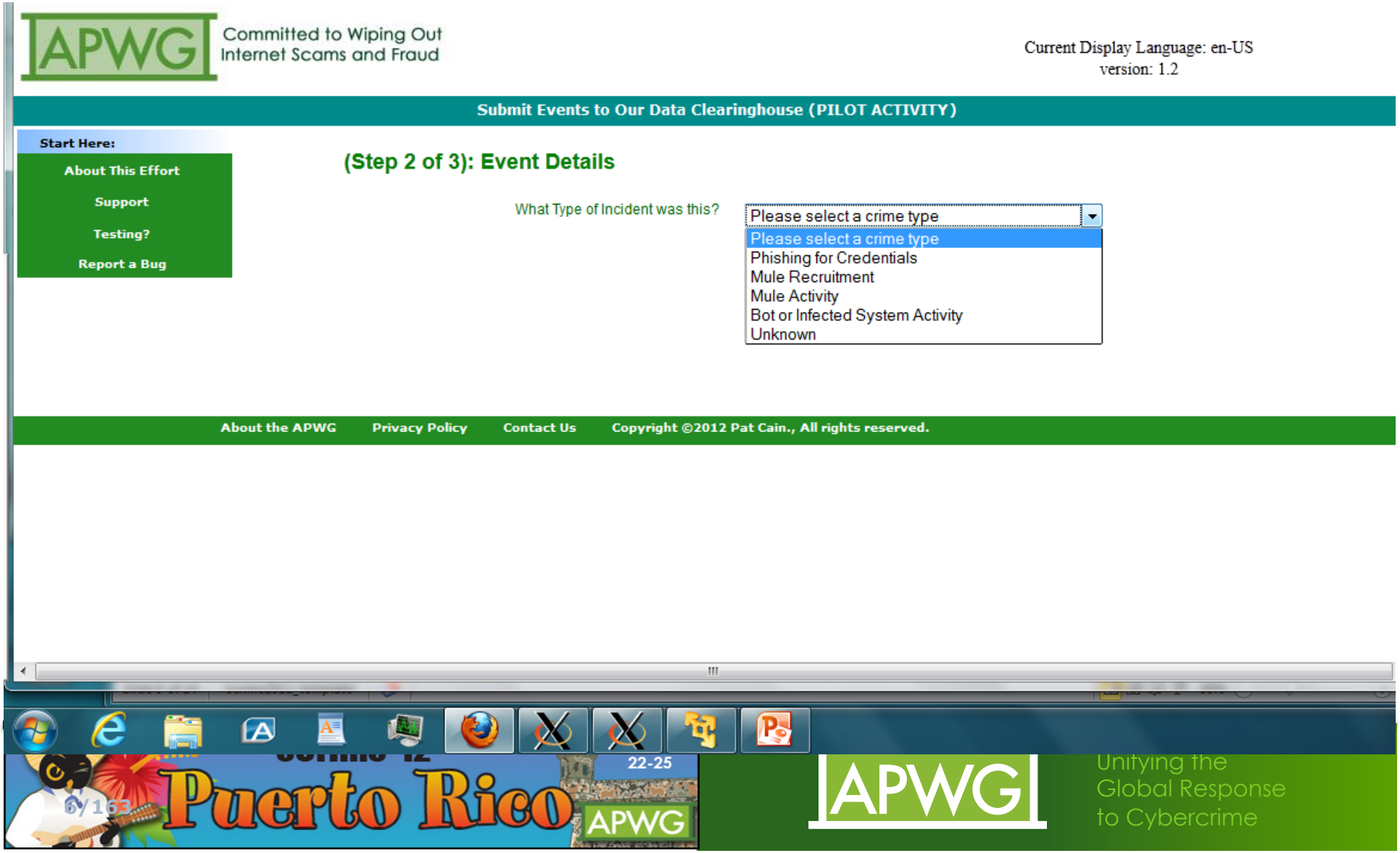
# Preparing for the Future

- As phishing changes you've asked for more data types
  - Phishkits
  - Infected systems
  - Malware distributors
  - Attacking systems (?)

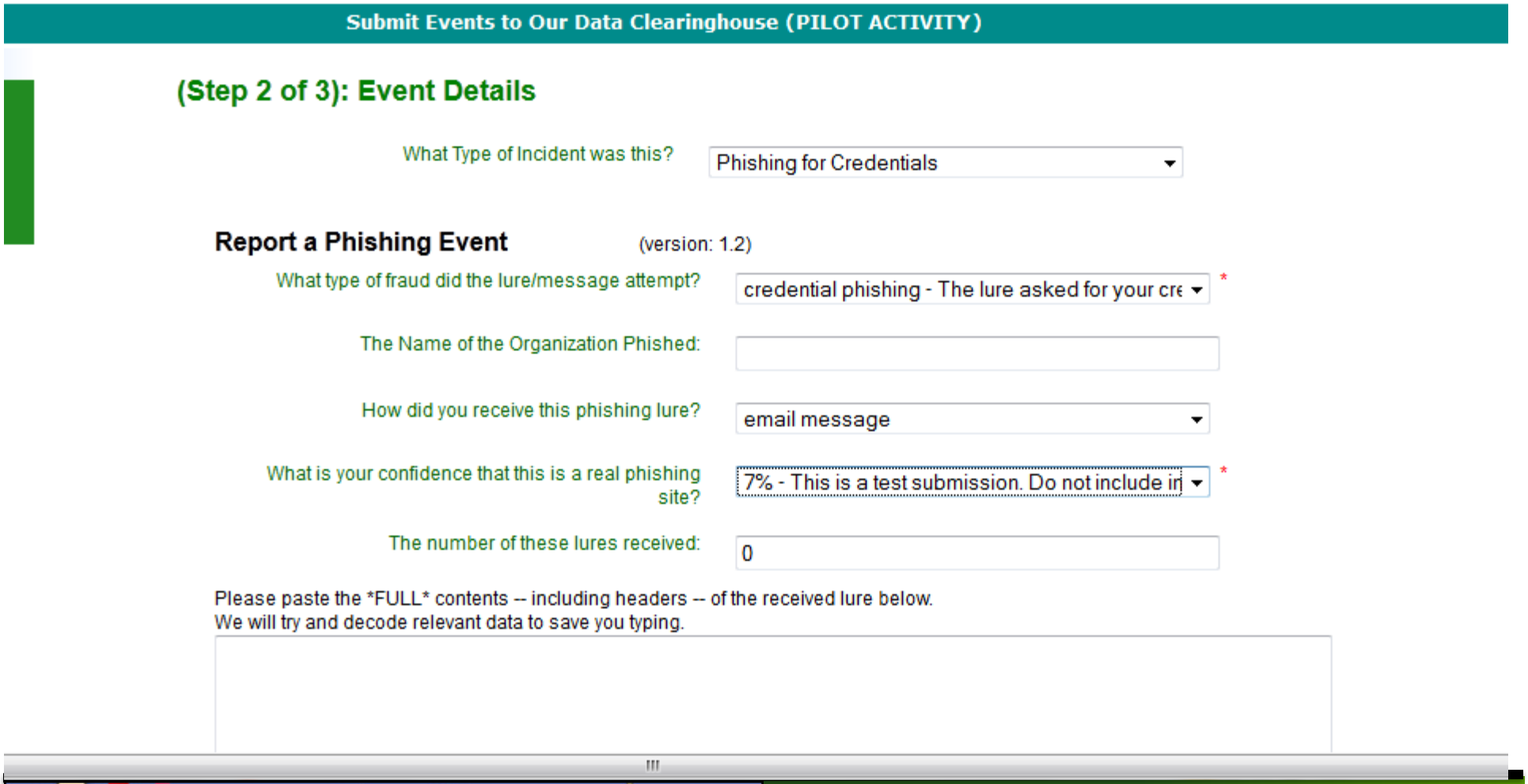
- Can we use the same processing for these new things?

# The webpage (draft)

# Data selector



# Phish page


Submit Events to Our Data Clearinghouse (PILOT ACTIVITY)
(Step 2 of 3): Event Details
What Type of Incident was this?
Phishing for Credentials
Report a Phishing Event
(version: 1.2)
What type of fraud did the lure/message attempt?
credential phishing - The lure asked for your cre
The Name of the Organization Phished:
How did you receive this phishing lure?
email message
What is your confidence that this is a real phishing site?
7% - This is a test submission. Do not include in
The number of these lures received:
0
Please paste the *FULL* contents -- including headers -- of the received lure below.
We will try and decode relevant data to save you typing.

# Bot Page

**Submit Events to Our Data Clearinghouse (PILOT ACTIVITY)**

## (Step 2 of 3): Event Details

What Type of Incident was this? Bot or Infected System Activity

### Report Bot Infection or Activity

(version 1.1)

What Date and Time was the Infection or Activity Seen? GMT Now

What is the Infected System's IP Address? * v4

What is the Infected System's DNS Name?

How confident are you of this infection? 85% - Experienced person looked at it *

The system is infected with what Bot or Infection (e.g., Zeus, Cutwail)?

This System was a member of which botnet?

The system was performing this activity (select or enter text) spam source

How was the infected system detected?

- Traffic Analysis
- Check-into C & C
- Contacted a Sinkhole

What was the Command and Control Systems's IP Address? v4

What is the Command and Control's DNS Name?

# Status

- The web pages are at  repoman.apwg.org/forms
- The scripts are on our ecrisp-x page
  - Sourceforge.net/projects/e-crispx
- There is a trac instance on repoman to report (the one last) bugs
- We currently accept phish and botz
  - Phish → UBL
  - Botz →
  - A search and retrieve interface is coming



# Related work: How to convey policy/sharing info

- Restriction markings
  - How to mark: Share with LEO? Friends? Public?
  - How to show: Know but no Touch!
- Can this data be shared with law enforcement?
  - There may be legal or other reasons involved
- Has the data aged out of sensitivity?
  - An IP with a bot in 1994 is not secret in 2012!
- We are piloting splitting the ‘marking’:
  - What is the sensitivity of the data
  - How we can share it

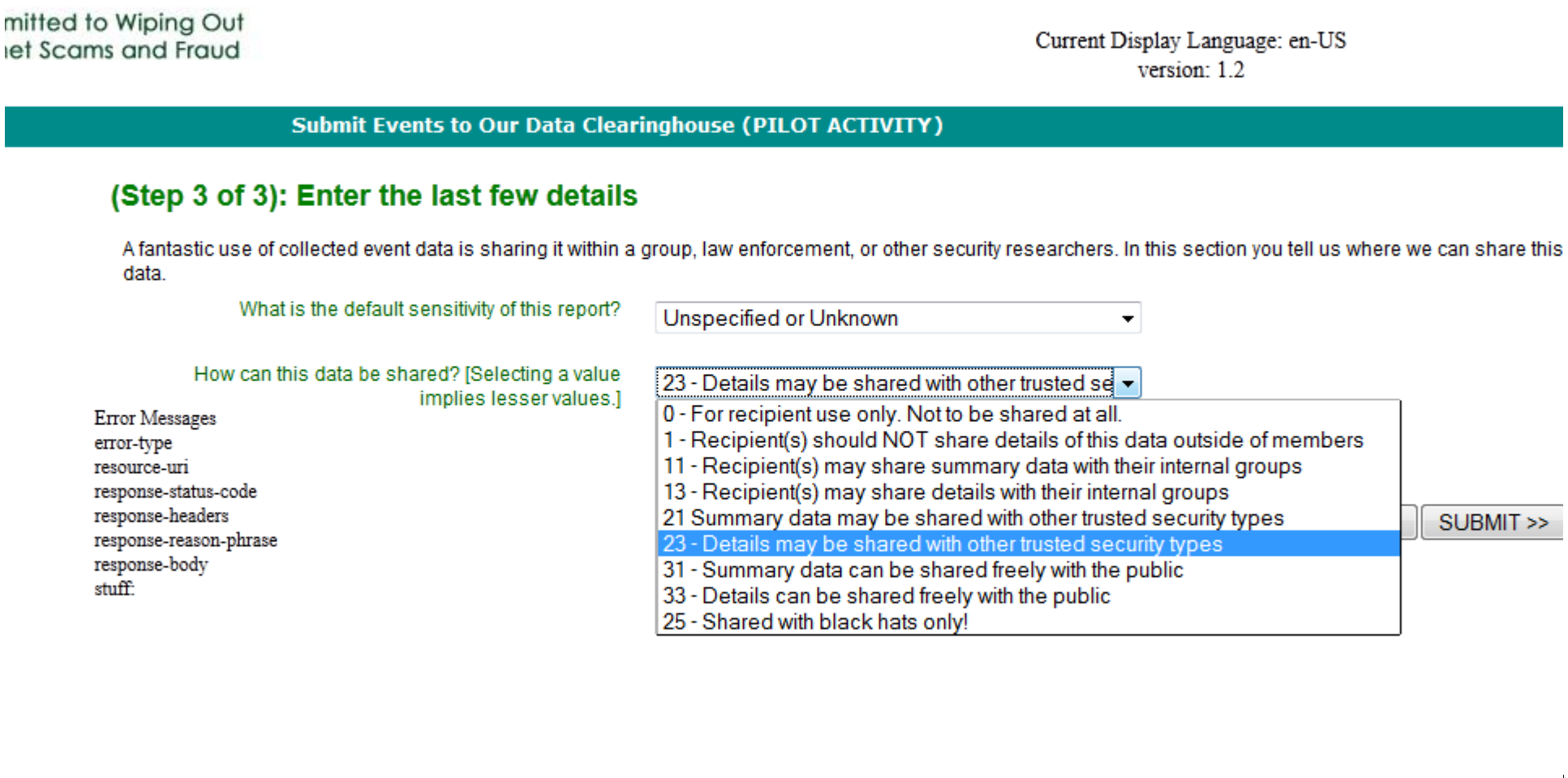
# silly and real examples

- Sharing with the 'Public':
  - 0 - Do not share
  - 1 - Summary data may be shared
  - 2 - Details may be shared
  - 3 - Too late. (already shared)
- Sensitivity marking:
  - Historical - Data is probably already known or is relatively old
  - Live-overt - Data was collected in a known fashion or collector
  - Live-covert - Data was collected via a private collection facility
  - Active - Data is currently part of an active LEO investigation
  - Active-no touch - Do not disturb or take actions on this dataset
  - Unspecified or Unknown

# An example…

- How can this data be shared within the APWG/ xxx?
  - 0 - For recipient use only. Not to be shared at all.
  - 1 - Recipient(s) should NOT share details of this data outside of group members
  - 11 - Recipient(s) may share summary data with their internal groups
  - 13 - Recipient(s) may share details with their internal groups
  - 21 - Summary data may be shared with other trusted security types
  - 23 - Details may be shared with other trusted security types
  - 31 - Summary data can be shared freely with the public
  - 33 - Details can be shared freely with the public
  - 25 - Shared with black hats only!

# Restrictions Pilot






Puerto Rico APWG

APWG Global Response to Cybercrime

# Open excitement

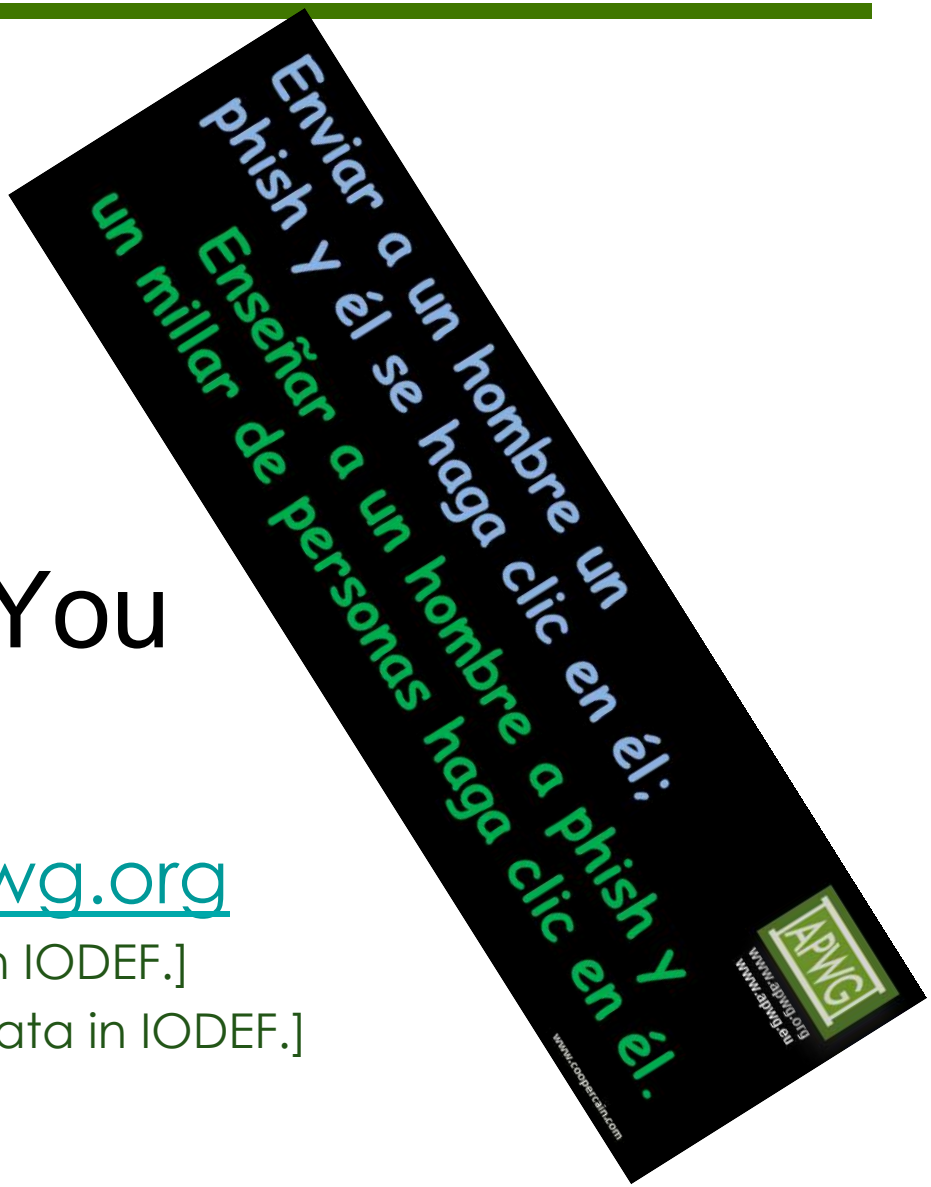
- Forms Integration into ECX
- How to integrate ECX data
  - Data from ECX entries into distributions
  - Script data into ECX
- Retrieval

# Your part…

- Feel free to submit something
  - There is a ‘confidence’ for “test” = not distributed
  - There is a  password on the web page
  - user=dataTester  pw=apwg2012  *
- The interface from the pages to the UBL is a little slow right now & the search is disabled.
- If you have ideas on how to give you the data, please advise.

- The repository maillist may become active again




# Thank You

pcain@apwg.org

[Send us data in IODEF.]

[Wait. Send us BOT data in IODEF.]