TAC Meeting
June 3, 2014

CFTC Data Landscape and
Update on Joint Efforts on Data Quality

Jorge Herrada, Office of Data and Technology

CFTC

Agenda

- Data at CFTC
  - Daily data volumes
  - Nine categories of data
  - Details of data categories
  - Data loading challenges
  - Trends in data volumes

- SDR Data Harmonization Update

Commodity Futures Trading Commission

Data Feeds
(approximate # of million **records** per day CFTC is currently loading)

Futures & Swaps Position Data
16M / Day

FERC Data
0.33M / Day

Futures Intraday Trades / Swaps Events
10M / Day (Note: adding order messages would be orders of magnitude greater, adding 200 – 300M records per day. At scale, adds another 7 - 10 feet to the diameter)

Market Data
0.2M / Day

Margin & Collat

Margin & Collateral
3M / Day

SDR Data (not to scale)

Valuation / Stress Test
0.1M / Day

Fin Statement / Capitalization
0.1M / Day

Org / Product Data
0.1M / Day

Time & Sales Futures Data (top of the order book)
240M / Day
(not to scale, at full scale this bubble would be 8 feet across)

2.5M records / inch

Enforcement
.2 – 1TB / Week
(not to scale, scale = 10 feet)

Nine Categories of Data

1. **Intraday Trades / Events Data** – All trades on futures exchanges and all events reported to SDRs

2. **Time & Sales Data** – Top of the order book at time futures trade executes.

3. **Position Data** - End-of-day long/short contract positions for futures and options or notional amounts for swaps.

4. **Market Data** - Real-time or end-of day market data (e.g. settlements, open interest, volume, bid/ask, last price, etc.)

5. **Margin & Collateral Data** - Daily or intraday initial margin, maintenance margin, variation margin, or collateral on deposit

6. **Valuation & Stress Test Data** - Daily data used to value futures or swaps positions (e.g. IRS curve/discount factor), or stress test results from what-if scenario analysis

7. **Financial Statement / Capitalization Data** - Daily or monthly data used from financial statement submissions

8. **Organization & Product Data** – Reference data about Organizations and Product specifications

9. **Enforcement Data** – Investigation and case files (including subpoenaed information and audio tapes)

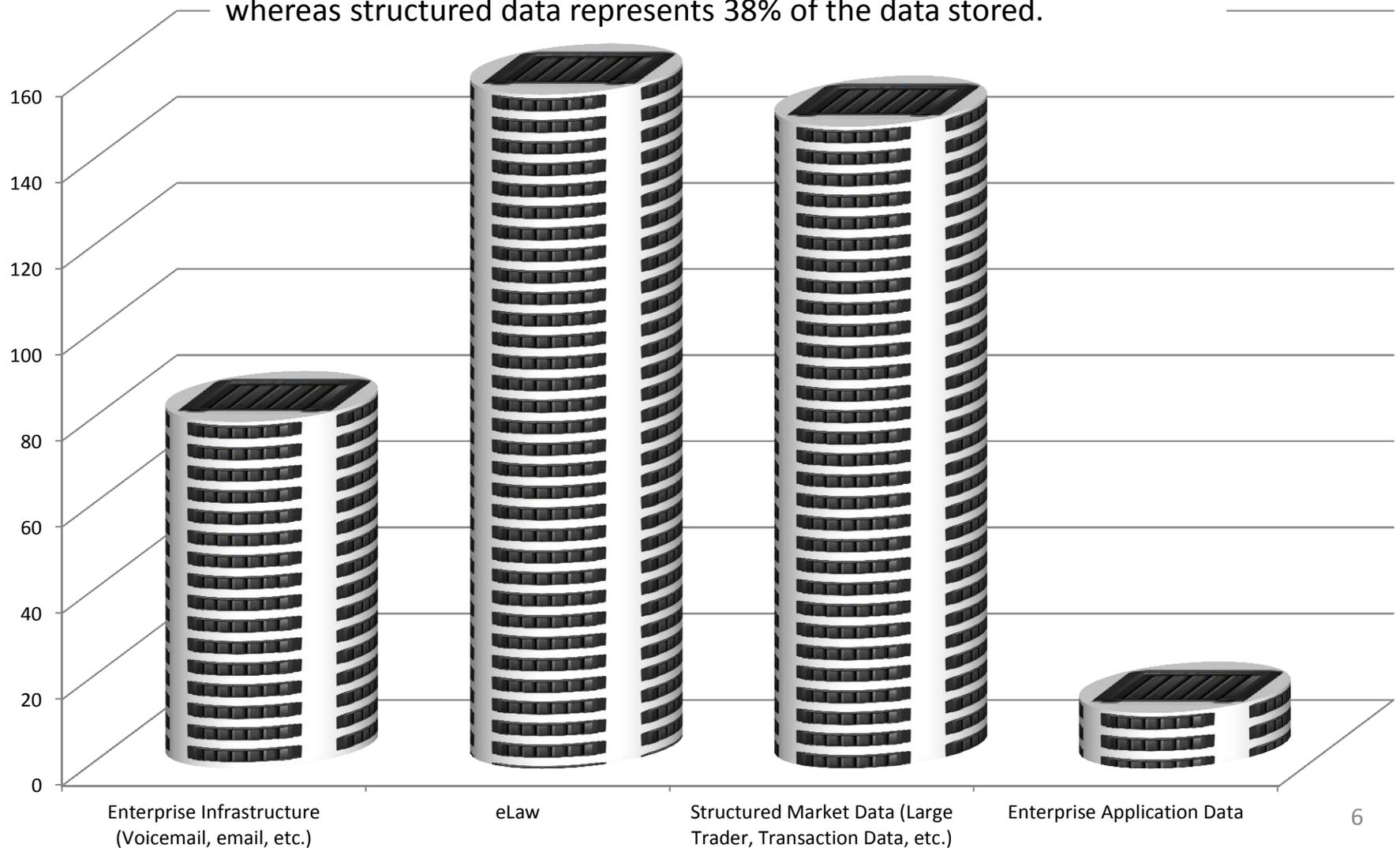# Commodity Futures Trading Commission

	Futures / Options		Swaps	
Intraday Trades / Events 8M	• Trade Data • CME Tag 50 • CME TCR Broker Ref	• CME Bulletin	• Part 45 (DDR only)	➢SDR Portals (DDR, ICE, CME, Bloomberg)
Time & Sales (order book) 240M	• Time & Sales Data • T&S Spreads			
Position Data 16M	• LT Pos Data (part 17) • Clearing Member Positions (part 16)	❖FERC LT Position export	• DCO data (part 30) • Part 45 (DDR only) • Part 20 Phys Commodities	• DCO Pos Data (Part 39) ➢SDR Portals (DDR, ICE, CME, Bloomberg)
Market Data .2M	• Market Data (part 16, price, volume, OI) ❖FERC market export	➢MDS – Reuters, Futures Source, Bloomberg	➢MDS – Reuters, Futures Source, Bloomberg	➢Markit
Margin & Collateral 3M	• CME Customer Margin		• DCO Acct (margin & collateral) • CDS margin & stats	• CME IRS Transfer receipts, margin, XV, positions.
Valuation / Stress Test .1M	• SPAN Margin, Risk Arrays	• ICE Non-Span Margin Engine	• IRS CME val., delta ladders, settlements, curves, & disc factors	• IRS LCH – delta ladders, margin, valuation ➢ ICC Margin API
Fin Stmt / Capital .1M	• Winjammer / RSR Loaders		• Fin Statement data • Monthly excess net capital	➢ NFA Daily Segregated funds
Org/ Product Data .1M	• Product Reference Files • Forms – DCO, DCM • Form OCR	• Form 102 (Spec Acct) • Form 40 • Form 204, 304	• Forms – SEF • Forms – SDR	➢ CICI / LEI (part 45) ➢ NFA Reg info ➢ Global LEI repository
Enforce-ment .2 – 1TB / wk	• Form – Tips Complaints & Referrals	• Subpoenas, massive data loads	• Forms – Tips, Complaints & Referrals	• Subpoenas, massive data loads

Commodity Futures Trading Commission
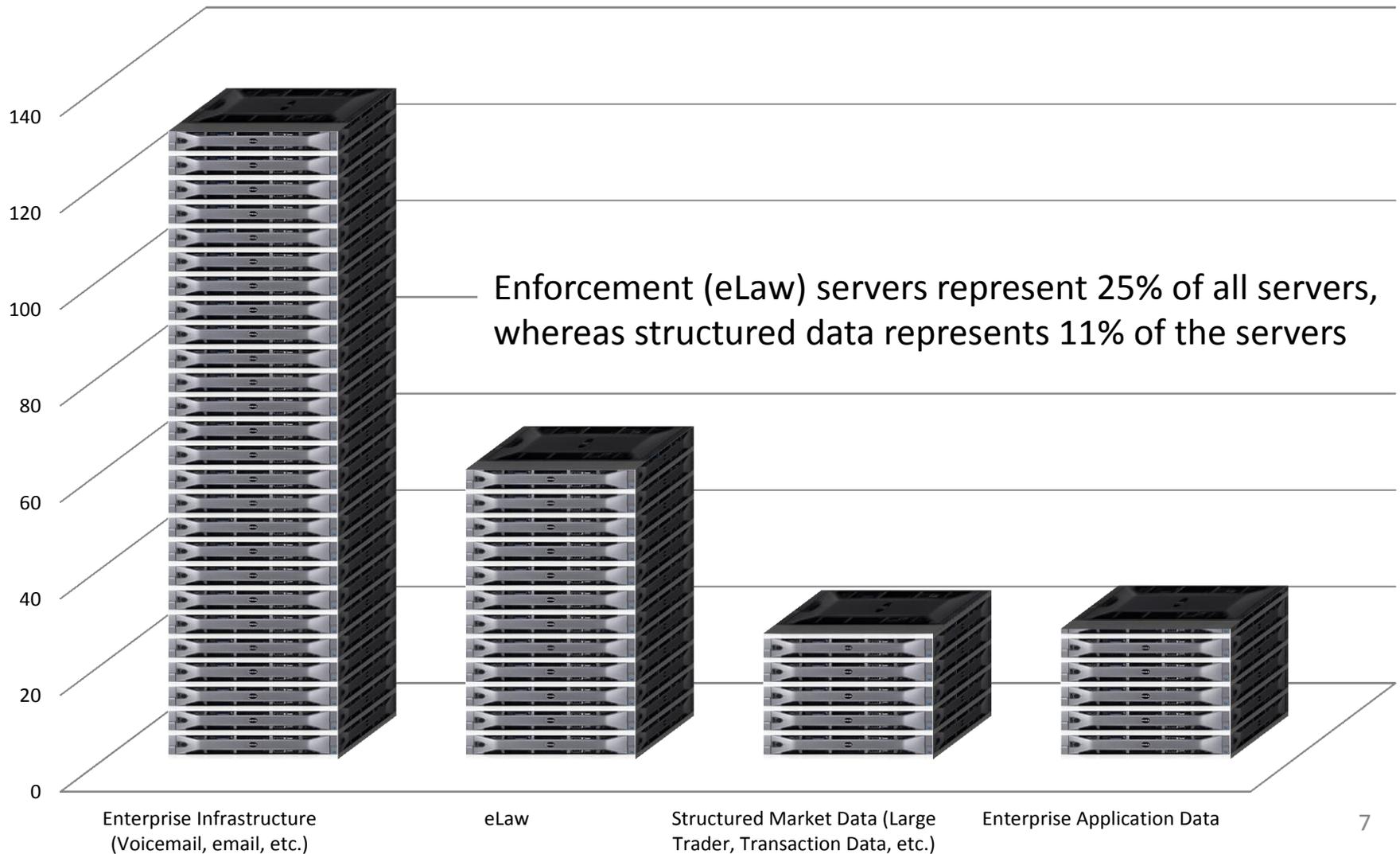
CFTC Network Storage Used by Program (TB)

Enforcement servers represent 39% of all data stored,
whereas structured data represents 38% of the data stored.

Bar chart. Y-axis: 0, 20, 40, 60, 80, 100, 120, 140, 160. Categories: Enterprise Infrastructure (Voicemail, email, etc.); eLaw; Structured Market Data (Large Trader, Transaction Data, etc.); Enterprise Application Data

6

CFTC Servers by Program

Enforcement (eLaw) servers represent 25% of all servers, whereas structured data represents 11% of the servers

Chart axis labels: 0, 20, 40, 60, 80, 100, 120, 140

Categories: Enterprise Infrastructure (Voicemail, email, etc.) | eLaw | Structured Market Data (Large Trader, Transaction Data, etc.) | Enterprise Application Data

Commodity Futures Trading Commission

Major eLaw Applications

(All of these apps in aggregate require a significant # of servers)
Highly standardized data, business functions and software products.

Case Tracking

Legal Hold Management

FOIA Tracking

Financial Statement Analysis

Digital Forensics

eDiscovery Processing

Trial Presentation

Audio Search

Legacy eDiscovery Review
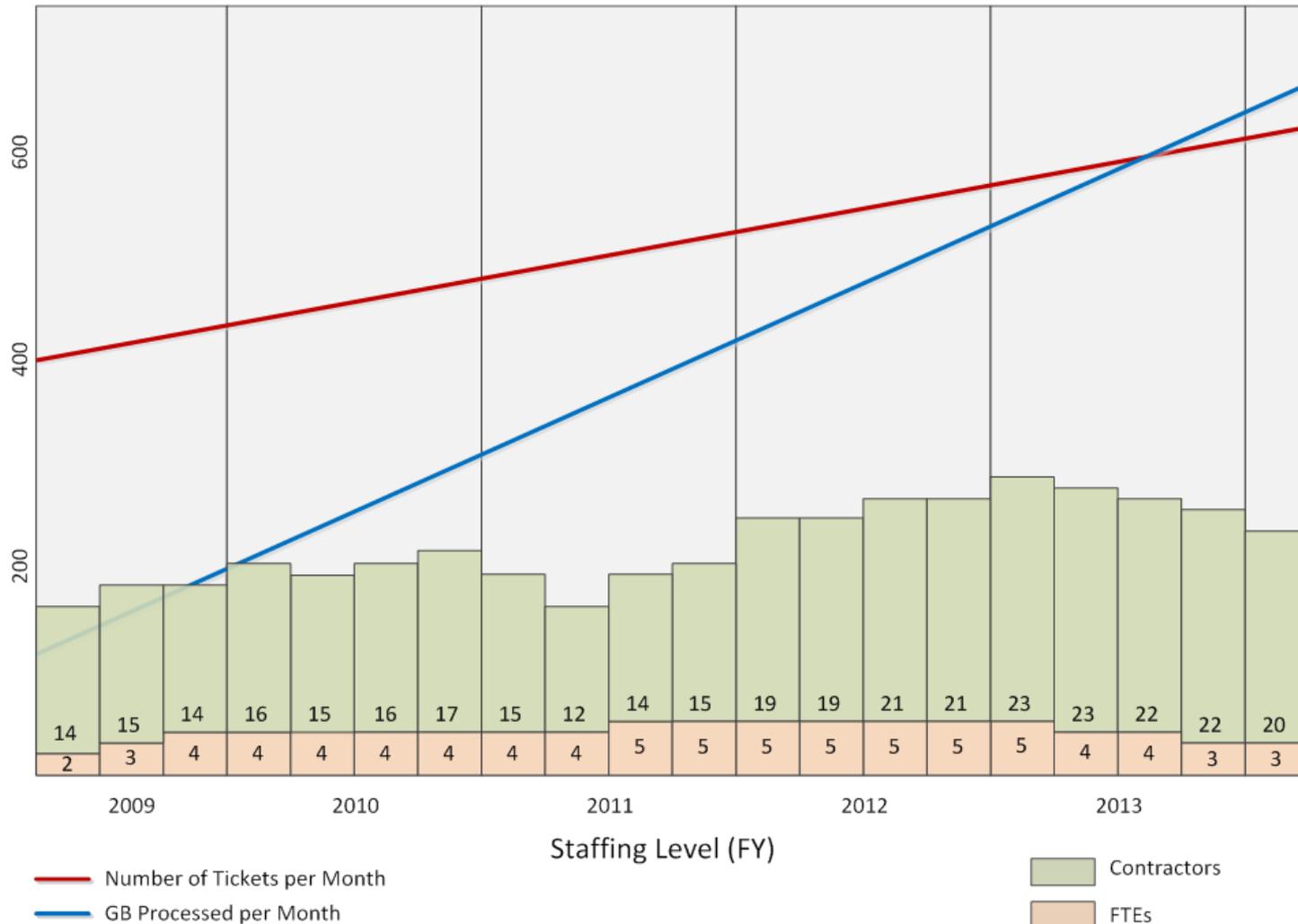
eDiscovery Review

Bubble sizes relative to importance to eLaw suite

eLaw staffing and data trends. Consistent software, data, and business functions are in place, so IT can ramp up data ingest with minor staffing implications.

Staffing Level (FY)

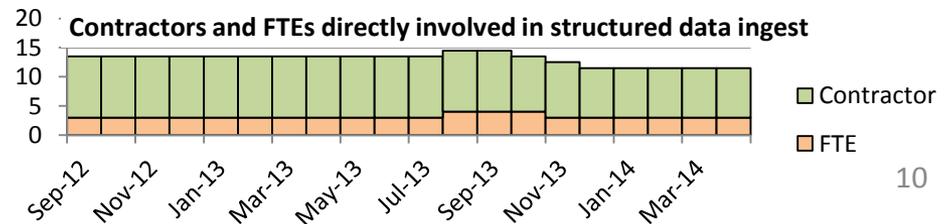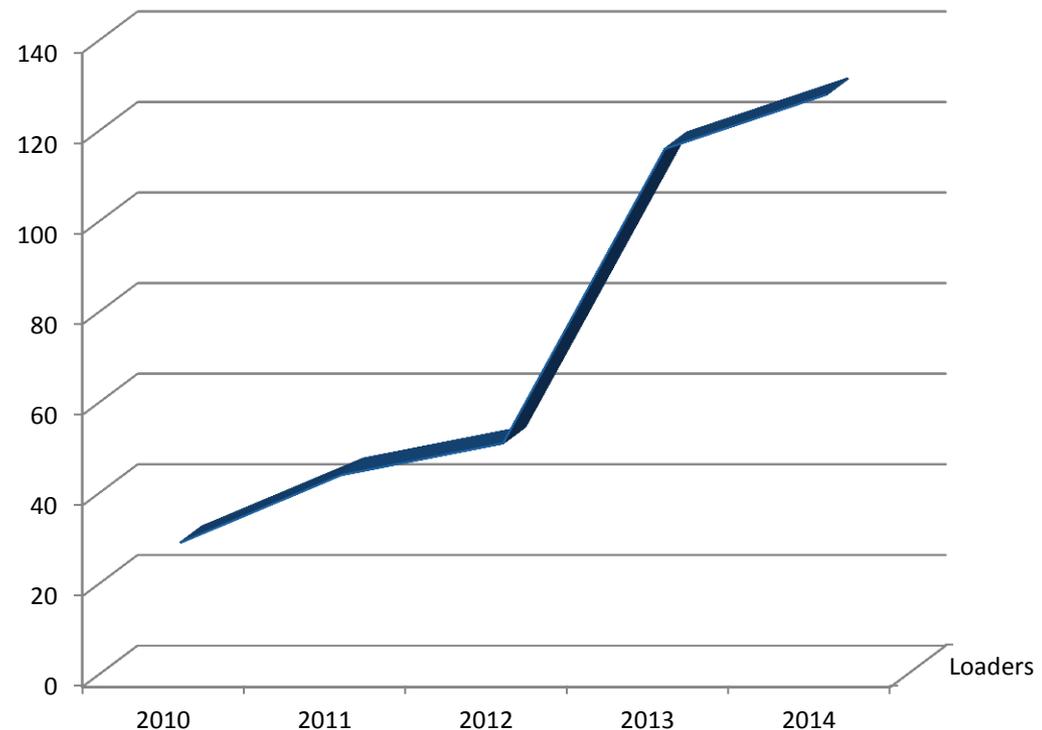| Legend	
Number of Tickets per Month	Contractors
GB Processed per Month	FTEs

9

Status of Structured Data Loading at CFTC

Increasing scope and complexity of structured data files (non-eLaw)

In 2013 we added loaders to handle 65 new types of data. Prior to 2013, we only handled 51 in total. In the last 12 months, we have added loaders for 77 new types of data. A 166% increase…*in just 12 months*

Each new data set requires data standardization, data architecture, custom loaders, data quality checks, monitoring, and feedback to submitters. Unlike eLaw, each new data set is a major project unto itself, analogous to standing up a new factory production line.

Data standards and harmonization are critical.

Types of Data Files

Contractors and FTEs directly involved in structured data ingest

10

Challenges in Loading Data at CFTC

Data Loading is now a 24hr per day operation

DATA RECEIPT TYPICAL SEQUENCE

Harmonization Update

- Phase I Harmonization implemented by SDRs for the credit asset class

- Phase II Harmonization (credit) in final stages of ODT review

- Scheduled bilateral meetings with SDRs for Phase II
  - Discuss progress on SDR harmonization action plans
  - Share Commission's independent data quality analyses with SDRs
  - Discuss exception reporting from SDRs to data submitters

- Plan a joint SDR meeting this summer for Phase II Harmonization

- Joint effort with Office of Financial Research (ORF)
  - OFR to assist with harmonizing IRS asset class
  - OFR to also assist Office of the Chief Economist with IRS research
  - IRS research to include swaps and futures data