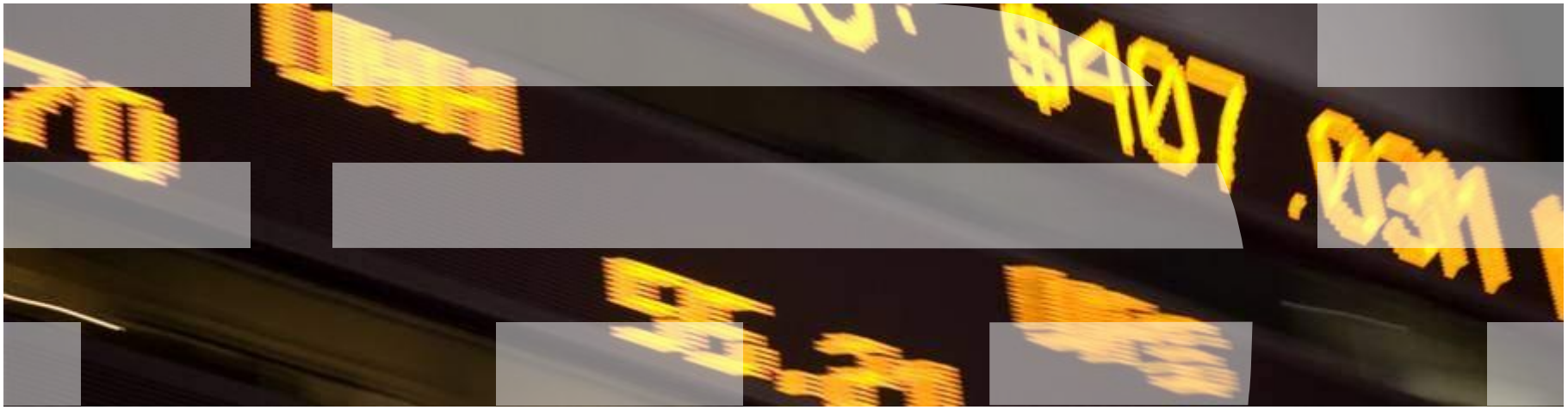
Showing Feasibility of fine grained Reference Data for Systemic Risk analytics



Francis Parr
IBM Research, July 2010

Data for Systemic Risk: Outline

1. Fine grained descriptive data on the financial system is key to detecting buildup of Systemic Risk
   - An instance where need for higher quality data is recognized

2. Shared data model capturing cashflows an important enabler
   - A novel need engaging/integrating existing sources of data

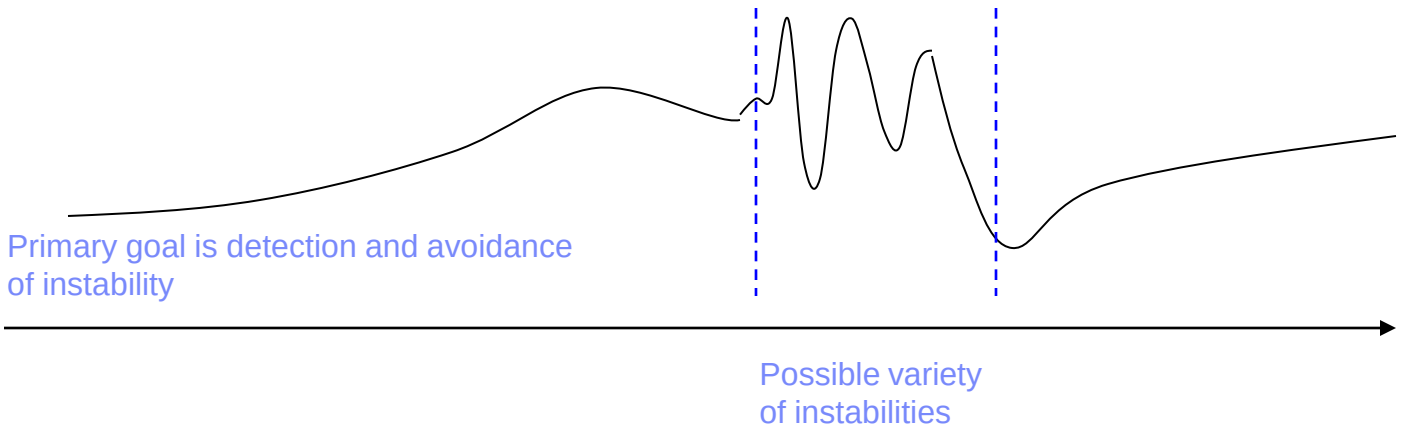
3. Systemic Risk Reference Data POC and feasibility demonstration

4. Completeness, security, privacy and openness questions

5. Summary: applying data quality concepts in an important, novel context

The value of fine-grained descriptive data

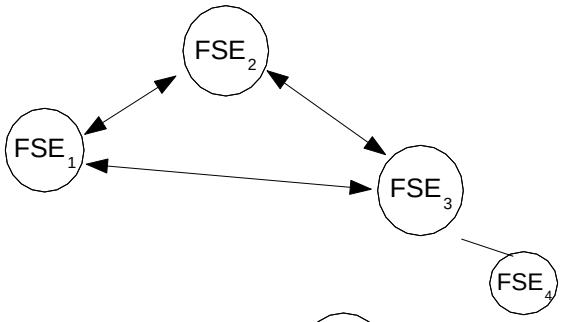
- **Analytics can be good predictors of behavior only when based on reliable system data**
- **Fine grained descriptive data on the state and history of the financial system is "objective"**
  - **Better placed to detect potential future crises than highly aggregated data**
- **Good descriptive data enables *multiple* analyses of risk**
  - **A portfolio risk detection approaches needed to avoid model risk**



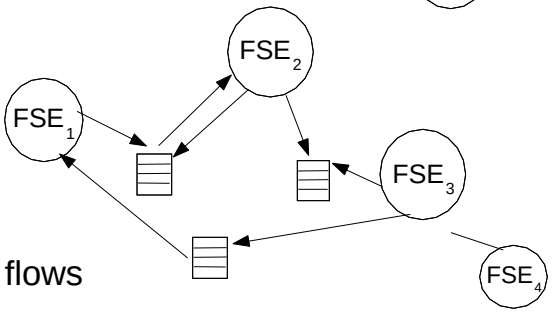
Highly aggregated vs. Fine grained models

FSE = Financial Services Entity

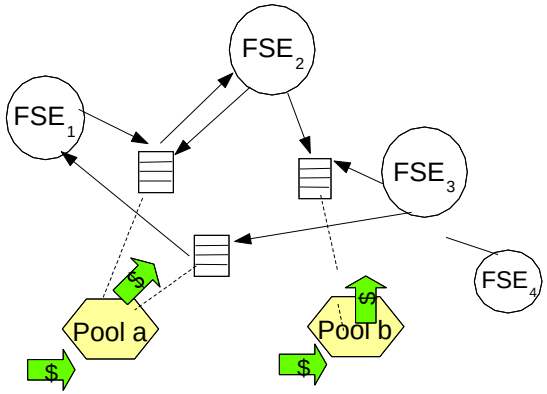
1. Network of FSEs with abstract dependency couplings
   - Model / predict viability dynamics
   - Ownership hierarchy



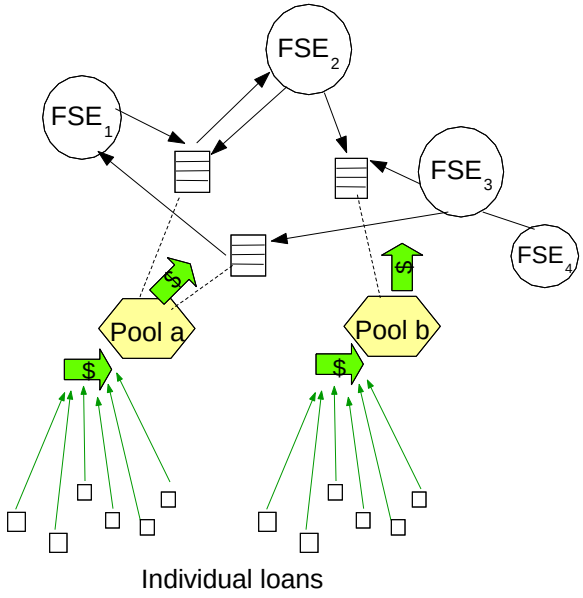
2. Network of FSEs with specified holder and guarantor dependency linkages (MBS)
   - Model /predict asset/liability flows



3. Network of FSEs with:
   - specified holder and guarantor dependency linkages (MBS)
   - Underlying pool and payment structure
   - Model / predict underlying pool cash flows (aggregated data)



4. Network of FSEs with:
   - specified holder and guarantor dependency linkages (MBS)
   - underlying pool and payment model
   - Model / predict individual mortgage default, prepayment behaviors



Gathering good fine-grained data requires…

- **Gathering high quality, fine-grained financial system data requires**
  - A system of systems architectural schema for the financial system
  - Data models and linkages capturing end-to-end cash flows
  - Unique identifiers for instruments, entities, assets etc
  - Standardized terms, glossaries, payout specifications – rigorous underlying semantics

- **A System of Systems description - cashflow coupled elements**
  - An end-to end “periodic table” view
  - Pluggable extensions within each column

- **Extends concept of publically available Reference Data**
  - Challenge in providing accurate linkages across domains
  - Ownership and identity relationships

- **Regulators will have access to information needed for this model**
  - Visibility to all participants in some domain

- **Standardized *semantic and data models* benefit many (all ?) players**

The Financial System as a "System of Systems" - data and analytics level (Periodic Table of Systemic Risk Models)

	I	II	III		IV		V
	Macro (real) Economy	**Micro economy (Financed Assets)**	**Finance System securitization supply chain**		**Finance System Liquidity (Trading, Hedging , Arbitrage)**		**Finance system Investment demand**
			Creators	*Instruments*	*FSE Actors*	*FSE State info*	
Scenarios	property values fall by 20% Unemployment rate raises	foreclosure rate in zip-code 12345 rises to 15%		equity tranche is wiped out	FSE xyz fails	Open FSEs ¦ Dark FSEs	Assets sold at loss. New capital needed to meet A/L ratios.
Models	Macro-economic model	Pre-payment model		Instrument Risk assess-ment model	FSE assessment models <how to model viability, contagion?>		Investment demand macro model
Data	Demographics Trade flows Interest rate Property Values (by zip code) Employment (by zip code) Consumption Investment	Mortgages Credit card Student loan Commercial Auto …	Ware-Houses Invest Bank securitize SIV,SPV	ABS RMBS CDO CDS Equities Bonds	Investment banks Trading broker /dealers Federal Reserve	Ownership hierarchy FSE assets / liabilities Hedging and Arbitrage strategies Liquidity assumptions	Insurance Companies Mutual funds Pension funds Invest bank Advisory Deposit Holder banks Hedge funds
	real economy		sell side		buy/sell		buy side

FSE = Financial Services Entity (firm)

Market based financial system

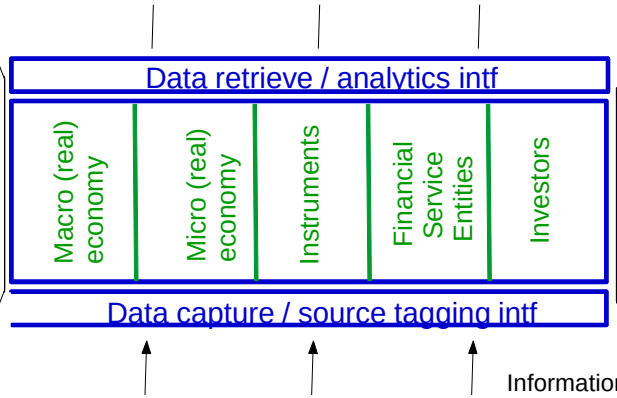
Semantic / data model and data store has central role in systemic risk

Multiple SR analytics algorithms

- Competing methodologies
- General vs special purpose
- Level of granularity
- Some published, most proprietary

Systemic risk analytics need

- Training and calibration data
- Many sources of data (broad scope)
- Consistently defined data
- Explicit cross domain linkages
- Affordable cleansing/validation costs



Semantic/data model and data store

- system of systems structure
- capture and analytics intfs

Systemic risk users can benefit from

- a standard data model
- Shared data stores / services for generally available data

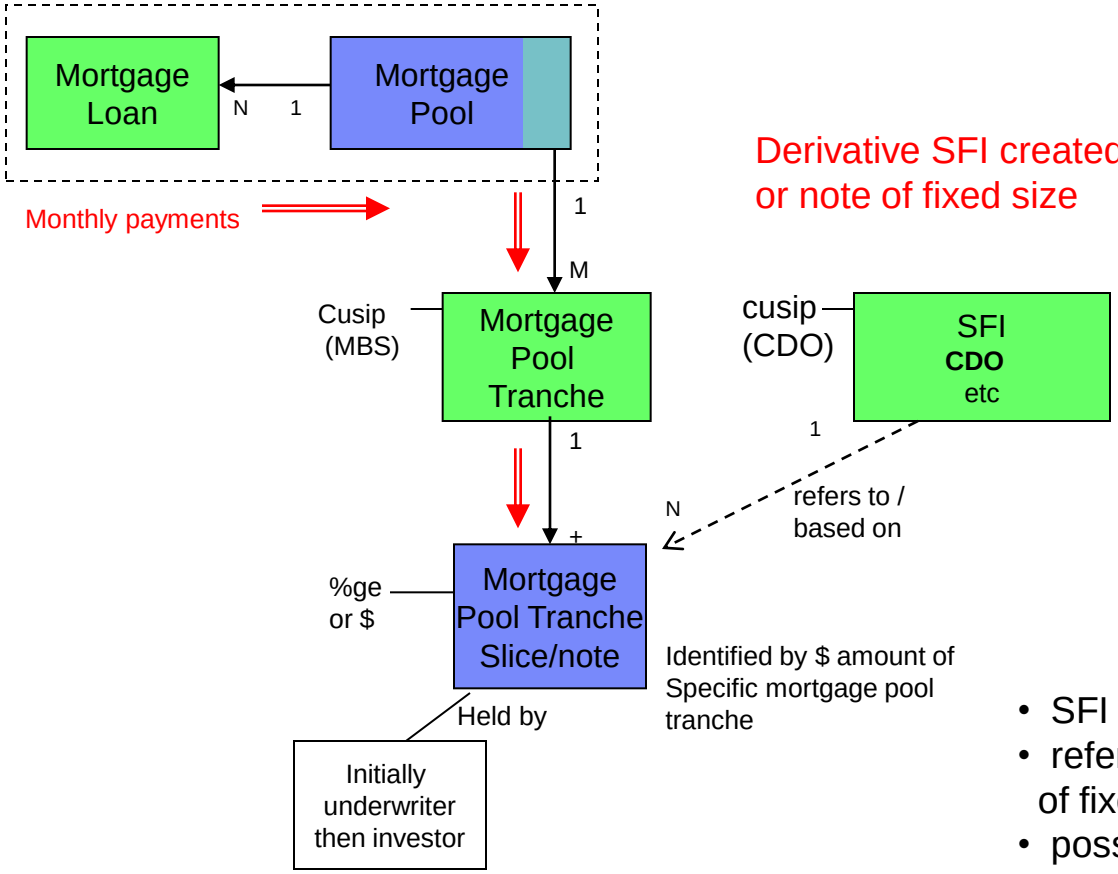
Multiple data sources

- different data domains
- different data suppliers
- some published, some is “for fee”
- generally available or private

The Systemic Risk Reference Data POC

- **POC** effort established June 2009 by IBM Research and Enterprise Data Management Council (Mike Atkin)
  - http://www.edmcouncil.org/default.aspx
- **Participation:** ECB, Federal Reserve Bank of NY, Freddie Mac, FHFA, Morgan Stanley, Swift, Standard and Poors, Interactive Data, Algo, GoldenSource, Moodys Analytics, MITRE, Polypaths, ….
  - Weekly calls
  - All work product in open shared team room
  - Develop open semantic model term and semantic data model assets
- **Goal:** Demonstrate feasibility of creating end-to-end, explicitly linked, fine grained data capturing mortgages, pools, deals, MBS, holder and guarantor FSE's using EDM semantics and glossary for base terms
- **Approach:**
  - Extend EDM semantic model to cover terms in selected space
  - Develop semantic data model reflecting data entities, their attributes and links
  - Validate semantic data model with feedback from
  - Generate implementation schema; populate with synthetic data – demo queries
  - Couple synthetic data – to participant risk analytics
  - Elaborate input side and data capture
- **Result:**
  - This shows that a regulatory requirement for source tagging by instrument issuers is feasible
  - It provides an example data model for a system risk facility

Overview of Mortgage Backed Securities (MBS) Model



Derivative SFI created referring to Mortgage pool slice or note of fixed size

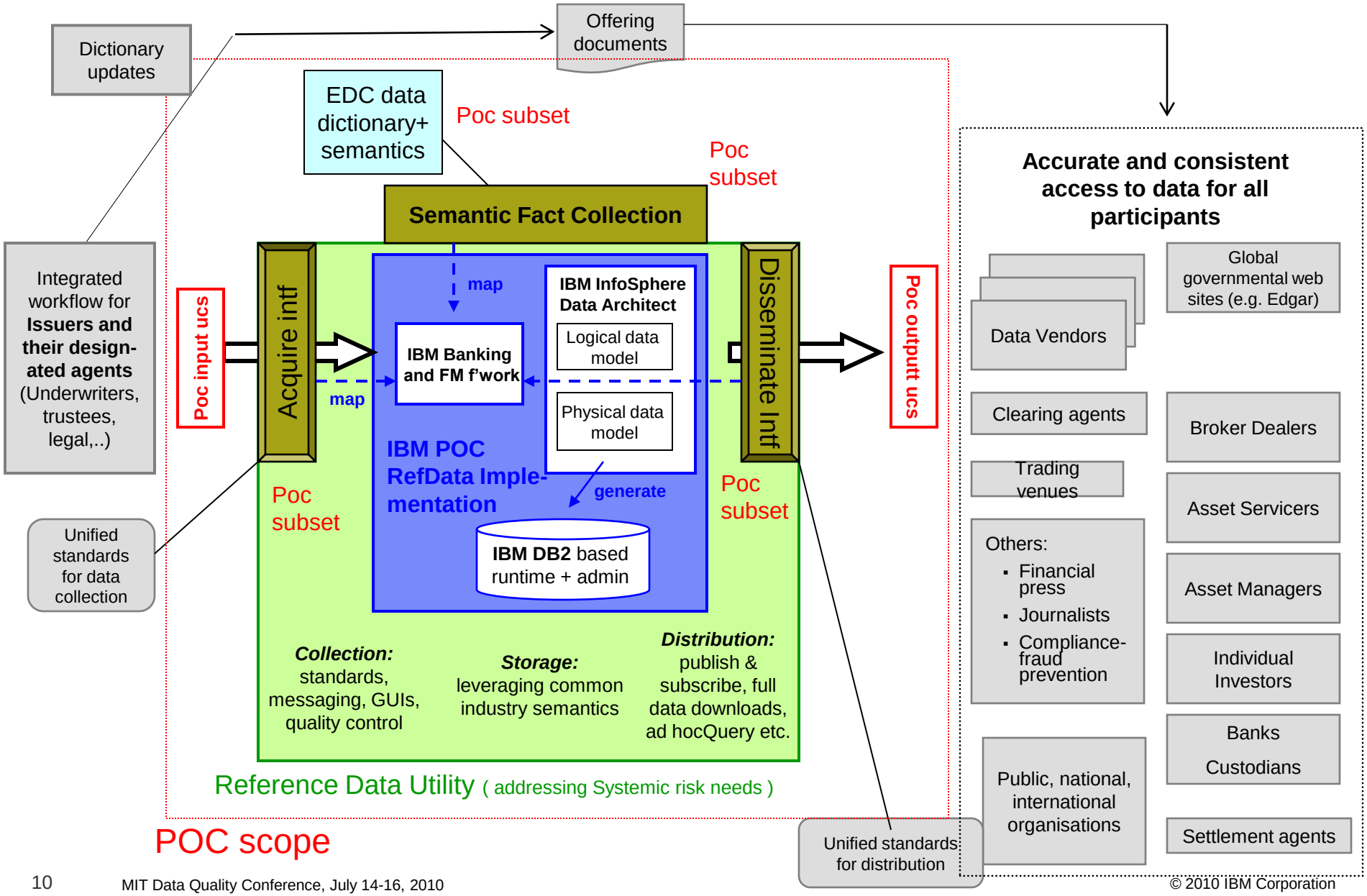
From data perspective,
missing reverse linkages

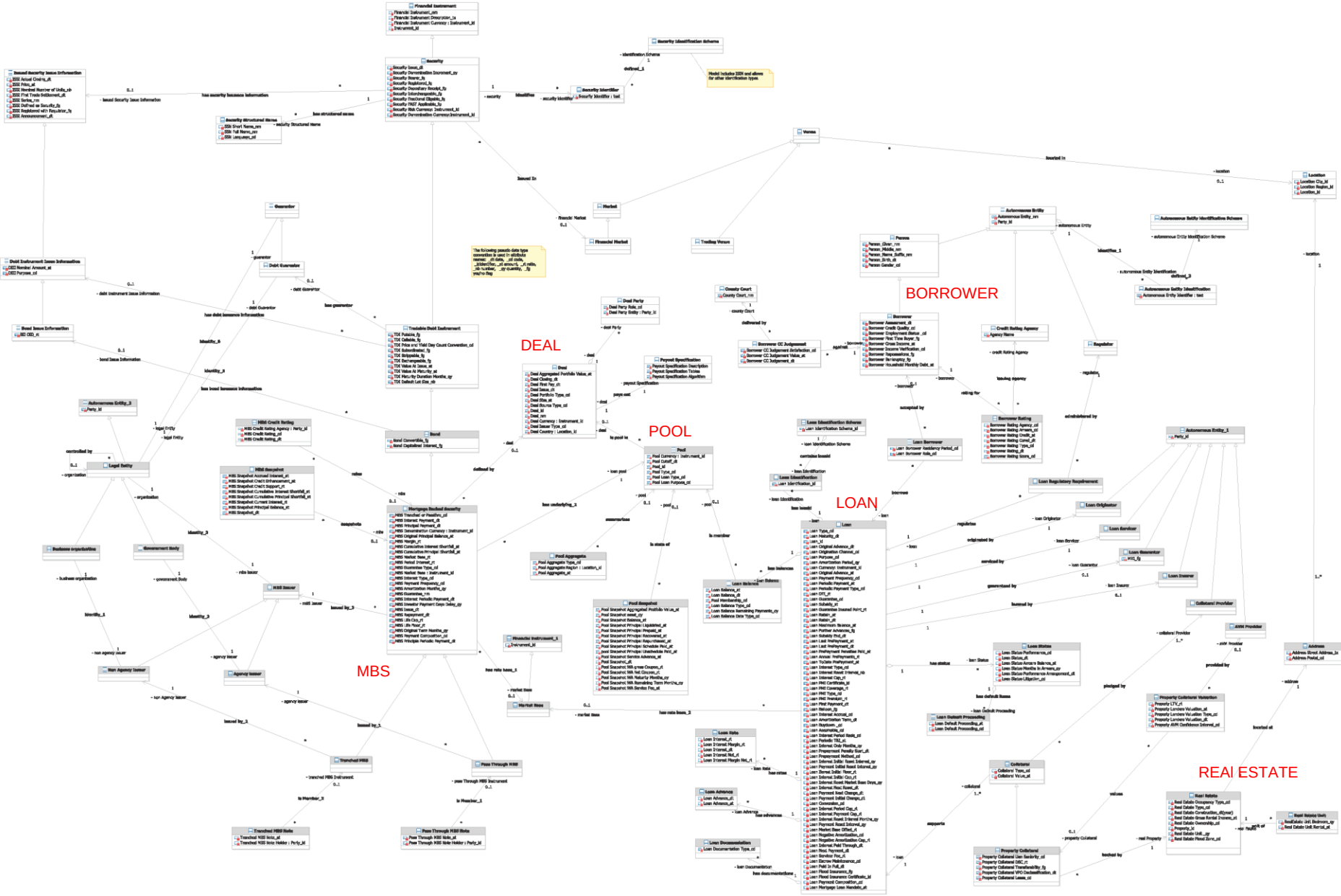
- Mortgage pool tranche -> mortgage loans missing

- SFI / CDO is a derivative
- refers to / based on some set of fixed size pool tranche slices
- possibly commitment to provide or buy.. a mortgage pool tranche slice .. Terms
- some combining function
- SFI is security with identifier / cusip
- many SFIs could refer to slices from the same mortgage pool

Overview of proposed POC test Data Flows and

Proof of Concept Reference Data Utility Architecture





Status in POC "Semantic fact" repository

1. **Generate physical data base schema for POC data model:**
   - Translation of POC model using IBM product tools and framework

2. **Automated generation of synthetic data to populate POC schema:**
   - Show reports to summarize and aggregate synthetic data
   - Show data drill down capabilities and inputs available for calibrating/training analytics
   - Synthetic data includes payment histories at MBS, pool and mortgage granularity level (since this information in principle reported publically)

3. **Show value of POC data in enabling predictive "what if" scenario assessment**
   - Gather sample "what if" scenarios to evaluate (with synthetic data) from POC work group participants
   - Demonstrate predictive systemic risk analytics for "what if" scenarios
     - With predictive analytics from POC members (IBM will provide one model set)
     - Multiple specialized and proprietary predictive models expected

4. **Work with data providers to replace synthetic with real data:**
   - Reuse same schema, reports, predictive analytics but now with real predictions
   - May involve data cleansing, transform etc
   - Mixtures of actual and synthetic data (improved distributions) possible
   - Possibly including access controls (as required for operational repository)

5. **Could evolve repository for other domains (beyond MBS) content**

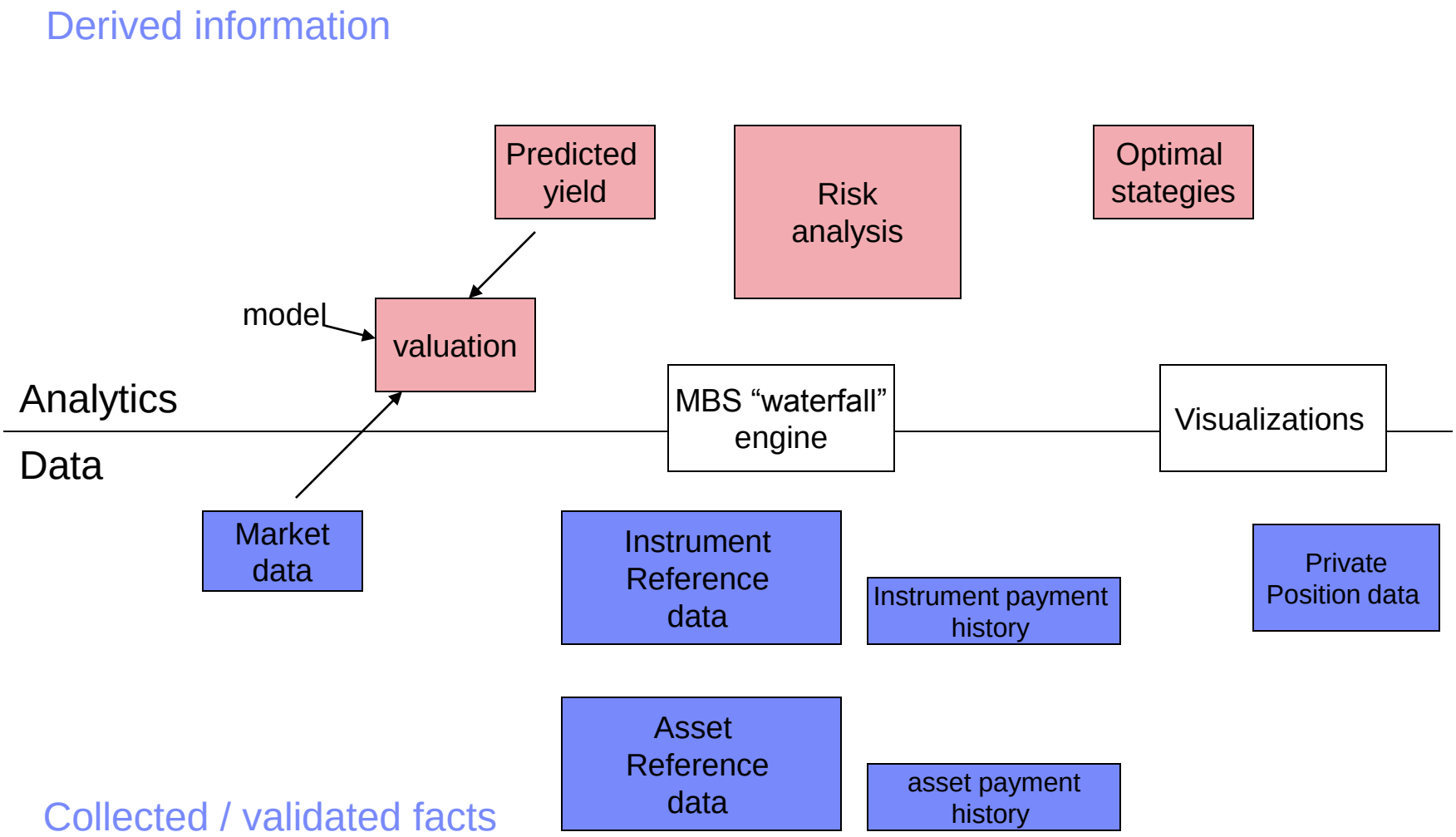
How complete is this data model?

- **Incremental – segment at a time -- approach necessary to capture behavior of complex financial system**
  - POC focus on Mortgages, pools, MBS, holders necessary
- **POC approach is extensible to other instrument and asset classes**
  - More complex CDOs,
  - More variable underlying asset classes, auto loans, commercial real estate
  - Full FSE ownership hierarchies
  - etc
- **General mathematical pay off function will be required for more complex instruments and derivatives**
  - Base MBS used pooling and tranching only
- **Model must be open ended and extensible**
  - Financial system will continually define novel classes of instrument and investment
  - Systemic risk regulators will capture new instruments as share of market becomes systemically significant
  - Continuous evolution
- **Approach must accommodate multiple regulatory domains**
  - Exchange of data with other regulators important
  - Clearly defined data model aids conversions

What data ? Who sees it? Who benefits?

- **Reference data on instruments is in principle public / generally available**
  - If combined with ratings, valuations etc often distributed for fee
  - A standard end-to end data model needed to combine reference date from multiple sources
- **Regulators will have access to private bank and FSE holdings data (from all FSEs)**
  - Must be held securely – access controls
  - Must be correctly linked to reference data
  - Provides a start point for estimating systemic risk
- **Banks and FSEs have**
  - Detailed knowledge of their own holdings
  - Access to public reference and market data
  - Estimates of counterparty holdings
- **Underlying asset histories (e.g mortgage payment) needs privacy protection**
  - Feasible through appropriate anonymization + aggregation
- **All above parties obtain benefit from standardized semantic and data model**
  - Combining generally available data from multiple sources with private data
  - Reduce costs and improve quality of generally available reference data
  - For fee information distributors benefit from growth of “broad scope risk analytics”
- **=> Successful Systemic Risk effort must build on interests of market players**

Analytics vs. Data – components in a risk system



Data for systemic Risk: summary

- **Transparency into financial systems at the transaction level is now practical**
  - Due to steady Moore's law decrease in processing and storage costs
- **Fine grained descriptive data enables multiple risk analysis and modeling approaches**
  - A portfolio of analyses is best approach to detect incipient systemic risk
  - Fine-grained more "objective" than aggregated data; better for detecting the "next crisis"
- **Gathering high quality fine grained financial system data requires**
  - A system of systems architectural schema for the financial system
  - Data models and linkages capturing end-to-end cash flows
  - Unique identifiers for instruments, entities, assets etc
  - Standardized terms, glossaries, semantics, payout specifications – rigorous underlying semantics
- **Systemic Risk Reference Data POC demonstrates feasibility in context of MBS**
  - End to end model relating: mortgages – pools – MBS – issuer/holder
  - Model validated in discussion with industry (POC) participants
  - Illustration subset populated with synthetic data – interfacing to analytics
- **Completeness, security, privacy and openness questions**
- **Conclusions and next steps (in the regulatory context):**
  - The POC approach can be generalized and applied to broader segments of financial system
  - Continuing work on capture / input gathering side and on coupling to analytics
  - Next step: combining data and stress testing in a systemic risk cloud
- **This is an example of adapting data quality concepts in a novel context**
  - Potential for legislatively driven document originator tagging

Backup Material

High Quality data is central to systemic risk analytics

- **A standard semantic and data model for financial system entities and their relationships is a key enabler for the growth and use of systemic risk analytics**
- **A single standardized data model will meet the analytic needs of most users and reduce data collection, cleansing and validation costs**
- **The semantic/data model connects**:
  Financial system reference, transaction and position data from many sources and domains

  Multiple systemic risk analytic algorithms (general / specialized, public / proprietary)
- **A "system of systems" component structure is used for the financial system model**
  For extensibility and flexibility – novel financial instruments will always be introduced

  Defines required explicit linkages of each entity to finer grained underlying entities and data
- **The semantic fact input interface and associated glossary**
  Enables consistent source document tagging and data entry by domain experts

  Enables coherent aggregation of data from multiple sources and data domains

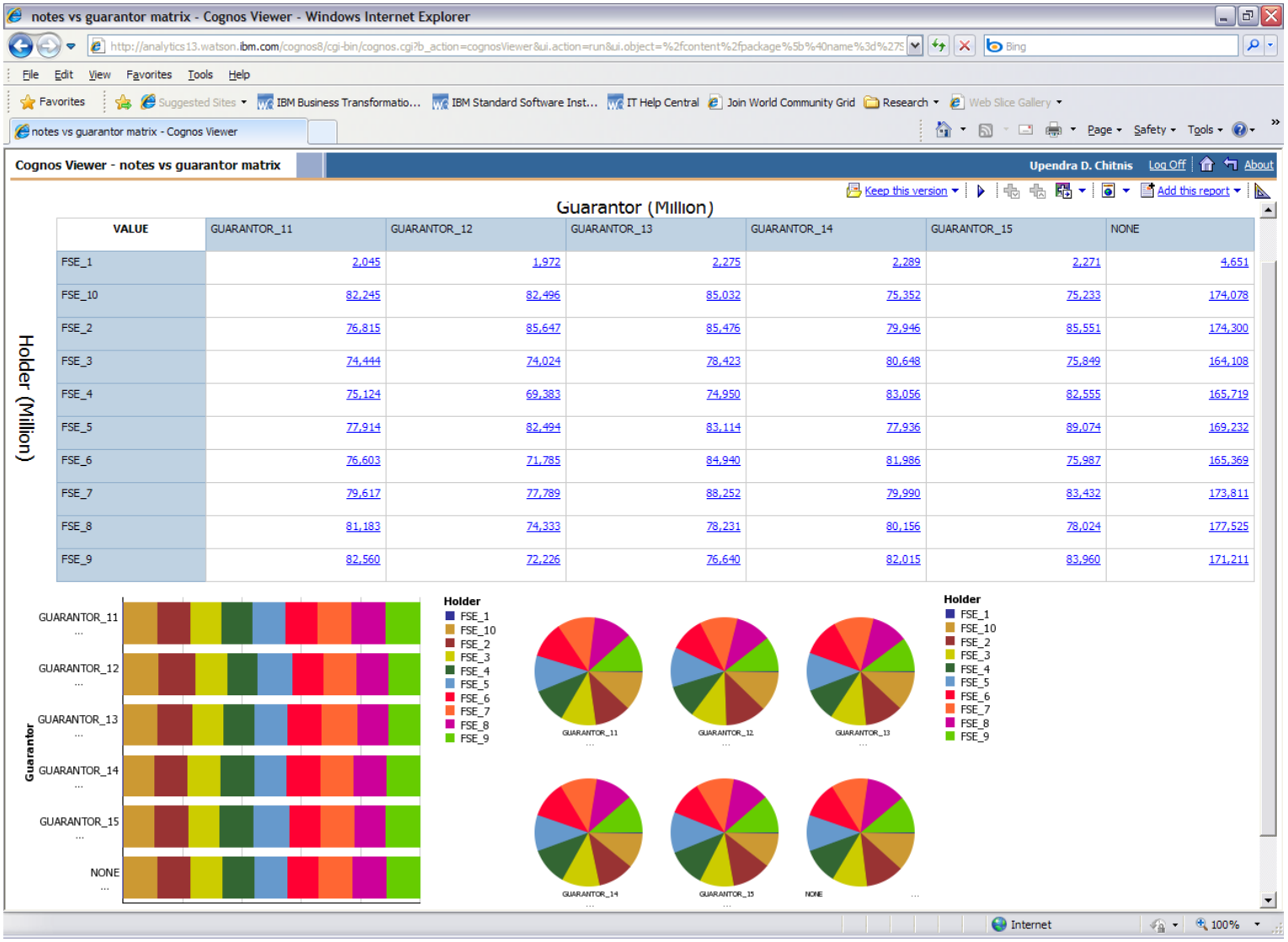
  Enables organizations to consistently extend generally available data with their private data
- **The query/retrieval interface allows a variety of system analytics to use and be calibrated from financial system data combining:**
  Generally available data (both for fee and published)

  Extended with restricted position data available to particular user organizations

Detecting Systemic Risk with High Performance Computing

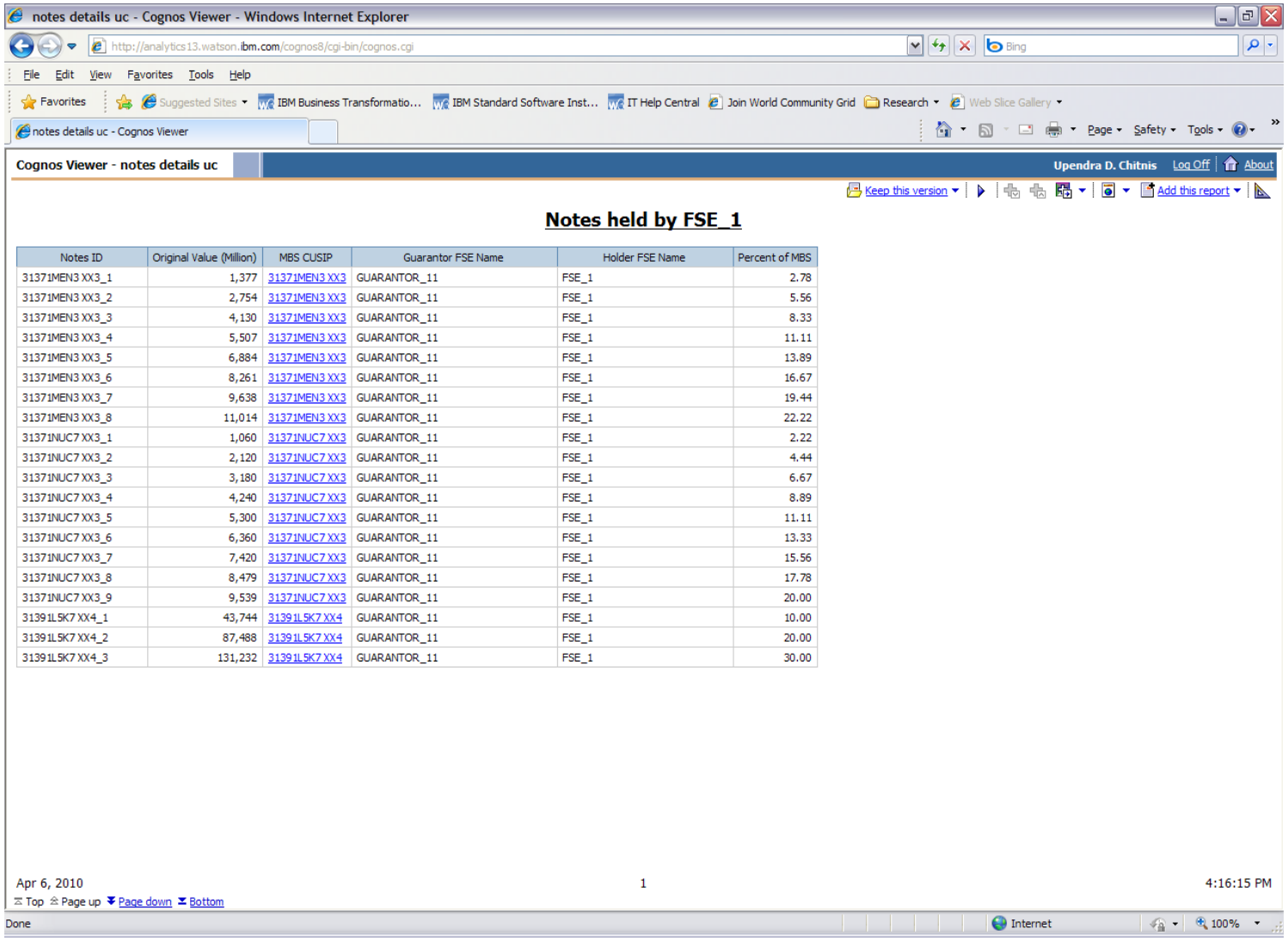
- **Financial Systems may be the most complex systems we deal with**
- **Traditionally, modeling and analysis of risk in Financial systems was attempted only with broad aggregate measures, (market indexes, weighted average coupons, …)**
- **With steadily falling processing and storage costs (Moore's law) …**
- **… it is now practical to capture the state and recent history of a financial system *at the transactional level* using High Performance Computing (HPC)**
- **Leading investment banks have been using intensive computing for trading risk, and, since the 2008 crisis, extending this for "broad scope" counterparty and liquidity risks**
- **European Systemic Risk Board, US Office of Financial Research establish need for regulatory understanding of Systemic Risk**
- **IBM Research launched a Systemic Risk Initiative in January 2009 to address this potential**
  - **Advance analytics relating to risk in complex systems**
  - **Financial system stability analysis as initial goal**
- **This talk is on "How to gather high quality data needed for HPC systemic risk analytics"**

Holder Guarantor Matrix



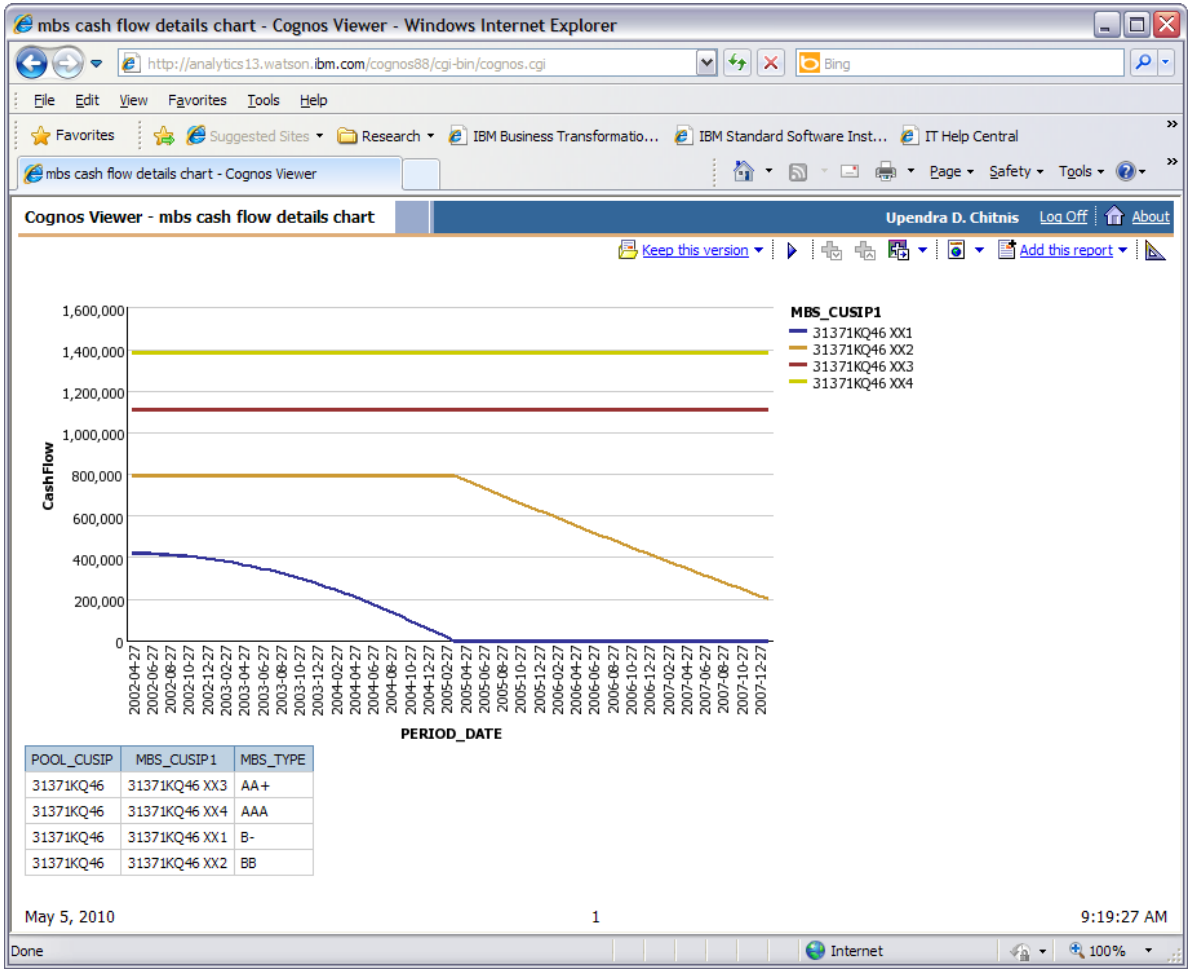
VALUE	GUARANTOR_11	GUARANTOR_12	GUARANTOR_13	GUARANTOR_14	GUARANTOR_15	NONE
FSE_1	2,045	1,972	2,275	2,289	2,271	4,651
FSE_10	82,245	82,496	85,032	75,352	75,233	174,078
FSE_2	76,815	85,647	85,476	79,946	85,551	174,300
FSE_3	74,444	74,024	78,423	80,648	75,849	164,108
FSE_4	75,124	69,383	74,950	83,056	82,555	165,719
FSE_5	77,914	82,494	83,114	77,936	89,074	169,232
FSE_6	76,603	71,785	84,940	81,986	75,987	165,369
FSE_7	79,617	77,789	88,252	79,990	83,432	173,811
FSE_8	81,183	74,333	78,231	80,156	78,024	177,525
FSE_9	82,560	72,226	76,640	82,015	83,960	171,211

Notes held by FSE 1

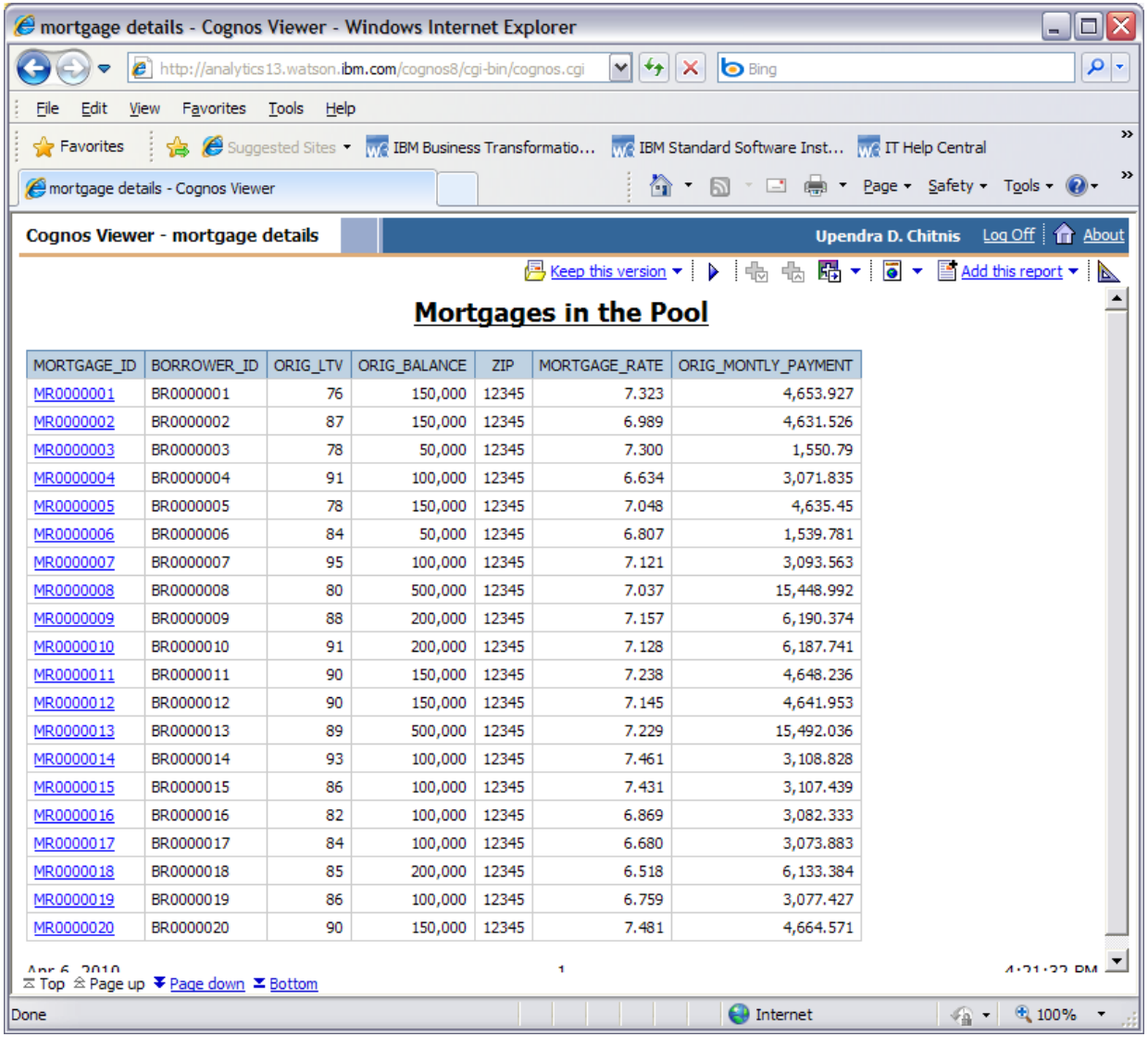


Notes ID	Original Value (Million)	MBS CUSIP	Guarantor FSE Name	Holder FSE Name	Percent of MBS
31371MEN3 XX3_1	1,377	31371MEN3 XX3	GUARANTOR_11	FSE_1	2.78
31371MEN3 XX3_2	2,754	31371MEN3 XX3	GUARANTOR_11	FSE_1	5.56
31371MEN3 XX3_3	4,130	31371MEN3 XX3	GUARANTOR_11	FSE_1	8.33
31371MEN3 XX3_4	5,507	31371MEN3 XX3	GUARANTOR_11	FSE_1	11.11
31371MEN3 XX3_5	6,884	31371MEN3 XX3	GUARANTOR_11	FSE_1	13.89
31371MEN3 XX3_6	8,261	31371MEN3 XX3	GUARANTOR_11	FSE_1	16.67
31371MEN3 XX3_7	9,638	31371MEN3 XX3	GUARANTOR_11	FSE_1	19.44
31371MEN3 XX3_8	11,014	31371MEN3 XX3	GUARANTOR_11	FSE_1	22.22
31371NUC7 XX3_1	1,060	31371NUC7 XX3	GUARANTOR_11	FSE_1	2.22
31371NUC7 XX3_2	2,120	31371NUC7 XX3	GUARANTOR_11	FSE_1	4.44
31371NUC7 XX3_3	3,180	31371NUC7 XX3	GUARANTOR_11	FSE_1	6.67
31371NUC7 XX3_4	4,240	31371NUC7 XX3	GUARANTOR_11	FSE_1	8.89
31371NUC7 XX3_5	5,300	31371NUC7 XX3	GUARANTOR_11	FSE_1	11.11
31371NUC7 XX3_6	6,360	31371NUC7 XX3	GUARANTOR_11	FSE_1	13.33
31371NUC7 XX3_7	7,420	31371NUC7 XX3	GUARANTOR_11	FSE_1	15.56
31371NUC7 XX3_8	8,479	31371NUC7 XX3	GUARANTOR_11	FSE_1	17.78
31371NUC7 XX3_9	9,539	31371NUC7 XX3	GUARANTOR_11	FSE_1	20.00
31391L5K7 XX4_1	43,744	31391L5K7 XX4	GUARANTOR_11	FSE_1	10.00
31391L5K7 XX4_2	87,488	31391L5K7 XX4	GUARANTOR_11	FSE_1	20.00
31391L5K7 XX4_3	131,232	31391L5K7 XX4	GUARANTOR_11	FSE_1	30.00

Principal Payment on MBS's in a single deal



Mortgages in the Pool



MORTGAGE_ID	BORROWER_ID	ORIG_LTV	ORIG_BALANCE	ZIP	MORTGAGE_RATE	ORIG_MONTLY_PAYMENT
MR0000001	BR0000001	76	150,000	12345	7.323	4,653.927
MR0000002	BR0000002	87	150,000	12345	6.989	4,631.526
MR0000003	BR0000003	78	50,000	12345	7.300	1,550.79
MR0000004	BR0000004	91	100,000	12345	6.634	3,071.835
MR0000005	BR0000005	78	150,000	12345	7.048	4,635.45
MR0000006	BR0000006	84	50,000	12345	6.807	1,539.781
MR0000007	BR0000007	95	100,000	12345	7.121	3,093.563
MR0000008	BR0000008	80	500,000	12345	7.037	15,448.992
MR0000009	BR0000009	88	200,000	12345	7.157	6,190.374
MR0000010	BR0000010	91	200,000	12345	7.128	6,187.741
MR0000011	BR0000011	90	150,000	12345	7.238	4,648.236
MR0000012	BR0000012	90	150,000	12345	7.145	4,641.953
MR0000013	BR0000013	89	500,000	12345	7.229	15,492.036
MR0000014	BR0000014	93	100,000	12345	7.461	3,108.828
MR0000015	BR0000015	86	100,000	12345	7.431	3,107.439
MR0000016	BR0000016	82	100,000	12345	6.869	3,082.333
MR0000017	BR0000017	84	100,000	12345	6.680	3,073.883
MR0000018	BR0000018	85	200,000	12345	6.518	6,133.384
MR0000019	BR0000019	86	100,000	12345	6.759	3,077.427
MR0000020	BR0000020	90	150,000	12345	7.481	4,664.571