GRUPPO TELECOM ITALIA

Seville 2014

PANEL: Information, Information, Information… How Much can we Handle?

ICCGI 2014
June 22 - 26, 2014 -
Seville, Spain

TELECOM ITALIA

PANEL: Information, Information, Information… How Much can we Handle?

PANELISTS

- Arno LEIST, Massey University, New Zealand
- Félix BUENDÍA, Universidad Politécnica de Valencia, Spain
- Jefferson CAPOVILLA, CPqD, Brazil
- Antonio J. TALLÓN-BALLESTEROS, University of Seville, Spain

Structure of the Panel:
Introduction on the topic
5 minutes speech per panelist

Discussion with the Audience is the key!

Who am I?

Claudio BOREAN

- **Leading SWARM – Beyond the Internet of Things Joint Open Lab of Telecom Italia**
- **Working on new innovative technical solutions leveraging on collaborative approaches among smart devices (Internet of Everything)**
- **Application fields: distributed energy systems, "swarm homes", collective intelligence for mobile applications**

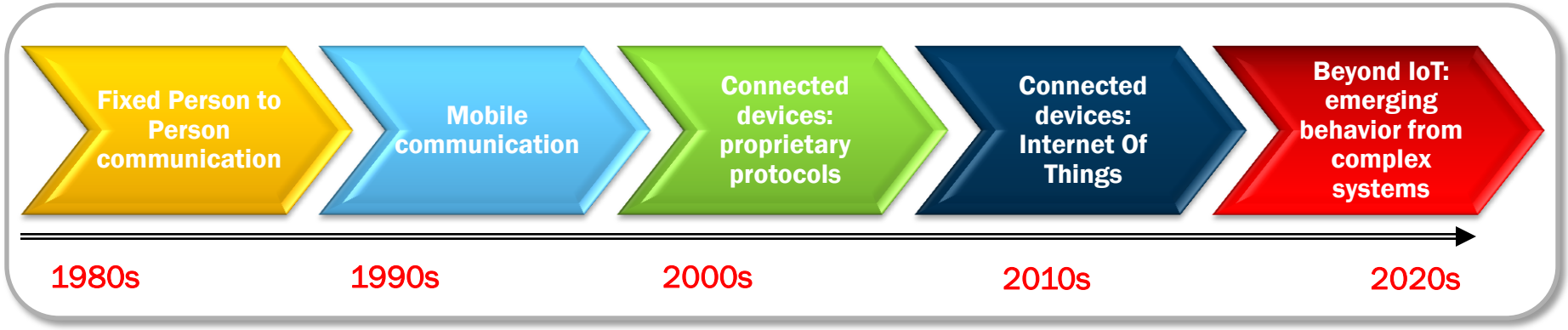
Areas: Swarm intelligence, Internet of Things, cooperative distributed applications

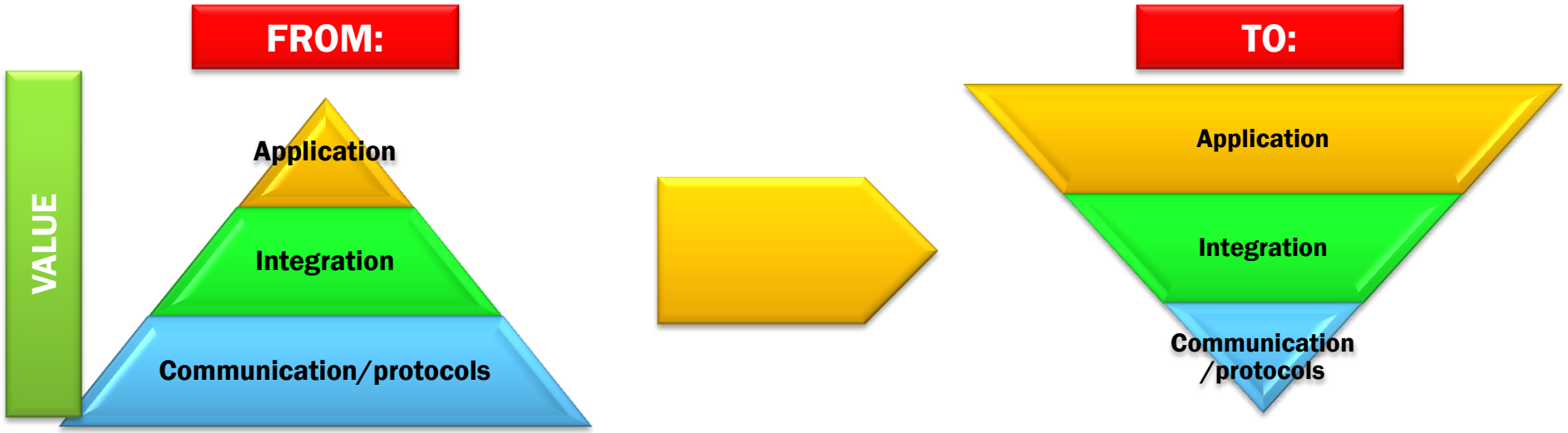
SWARM

Beyond the Internet of Things

http://jol.telecomitalia.com/jolswarm/

Topic summary: From “connected devices” to “beyond IoT” and “IoE”





PRESENTATIONS FROM PANELIST

DISCUSSION

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

InfoWare 2014: ICCGI Panel

Information, Information, Information... How Much Can We Handle?

Arno Leist, PhD
a.leist@massey.ac.nz

Who Am I?

- Lecturer in Computer Science
- Massey University, Auckland
- Research interests
  - Parallel computing
  - Data-parallel accelerators
  - Computational simulations
  - Visualisation

Processing Data

- A computational challenge
  - Processors are not getting much faster any more
  - But they are getting more parallel
- All components of the system must scale with the problem size
  - Processors
  - Memory bandwidth
  - Network bandwidth
  - Algorithms (limited by Amdahl's law)
  - Parallel frameworks

Processing Data

- A computational challenge
  - Processors are not getting much faster any more
  - But they are getting more parallel
- All components of the system must scale with the problem size
  - Processors
  - Memory bandwidth
  - Network bandwidth
  - Algorithms (limited by Amdahl's law)
  - Parallel frameworks
- What we call *big data* today is just data tomorrow

What To Use?

- Which architecture and environment should your software target?
  - x86, ARM, Power, ...
  - Embedded, mobile, PC, servers, compute clusters, supercomputers
  - Accelerators
  - Cloud computing
- Which languages / frameworks should you use?
  - Programming languages: C++, Java, Go, Python, CUDA, OpenCL, ...
  - Multi-threading libraries: PThreads, OpenMP, …
  - Multi-tasking libraries: TBB, Cilk Plus, ...
  - Data-parallelism: intrinsic instructions, …
  - Distributed computing: MPI, Hadoop, …
- Where and how should you store your data?

Discussion

- Where do you see parallel computing going in the next 5 to 10 years?
- What is keeping *you* from utilising all the parallelism available in your computer?
- Do we really have to process *all* that data?

Mobile Technologies

New chances for education

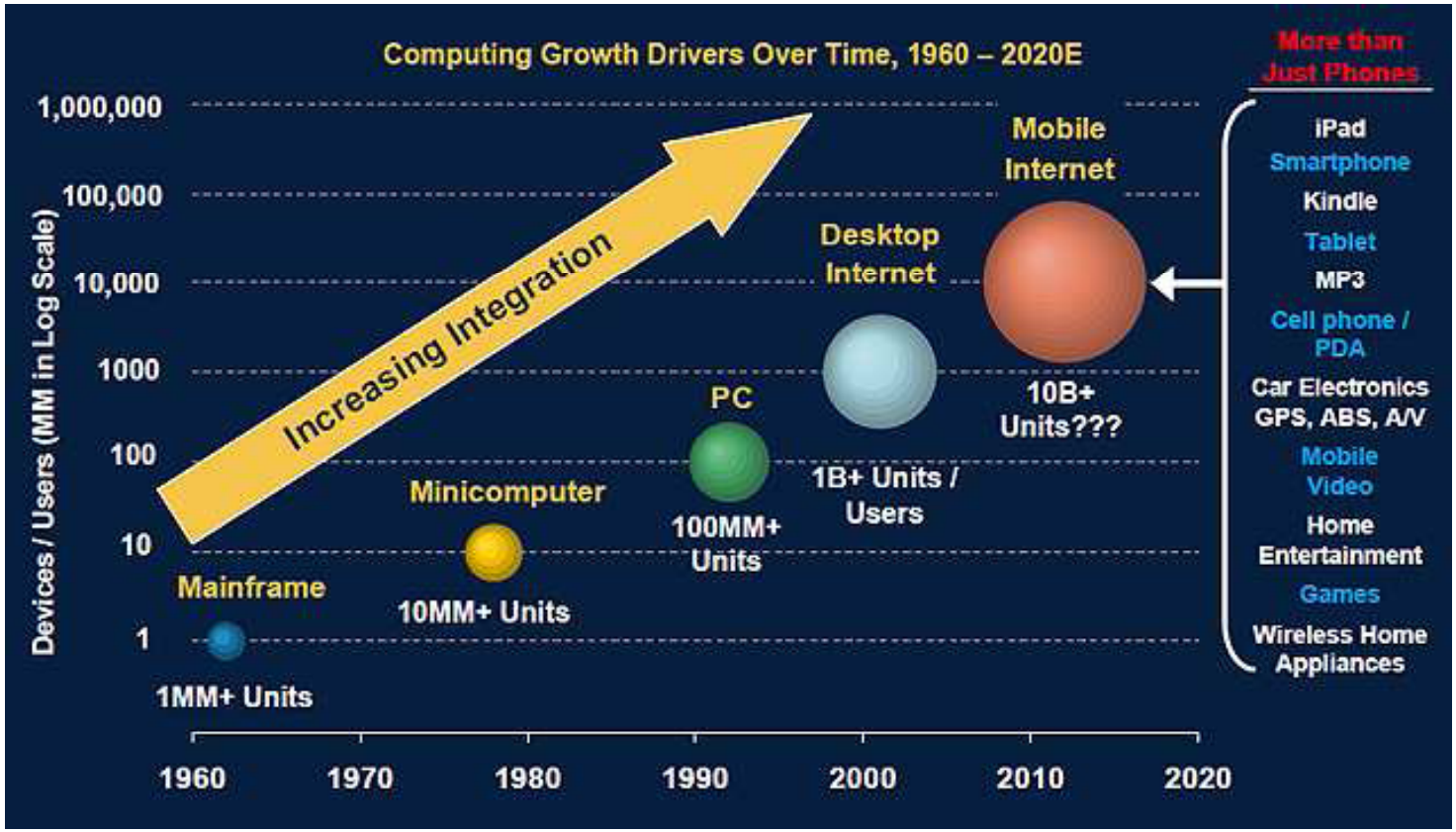
Félix Buendía

Ninth International Multi-Conference on Computing in the Global Information Technology
ICCGI 2014
June 22 - 26, 2014 - Seville, Spain

PAID-UPV/2791

Mobile technologies

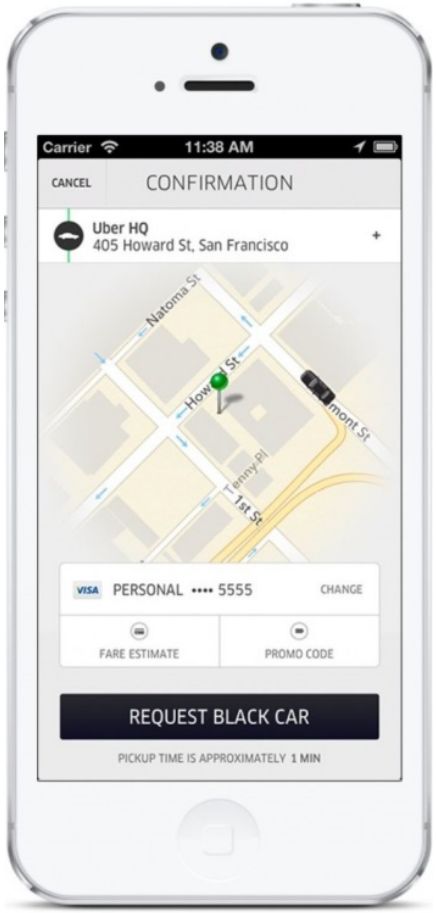
- Evolution towards Mobile Internet



www.morganstanley.com/techresearch

Mobile applications

- Moving to mobile apps



http://blog.uber.com/2012/04/18/chicago-taxi-uber/

Mobile education

- Platforms & environments

Mobile education

- Current status
  - Mobile Training Implementation Framework (MoTIF)

http://www.adlnet.gov/mobile-learning/motif/

Mobile education

- Challenges & opportunities
  - Need to address
    - supporting alternative learning/instructional methods
    - leveraging the capabilities of the mobile platform
  - Some questions
    - Is enough converting existing eLearning courses /MOOC by only “resizing” them to smaller screens?
    - What services and functionalities are required in the mobile learning context?

Information, Information, Information… How Much can we Handle?

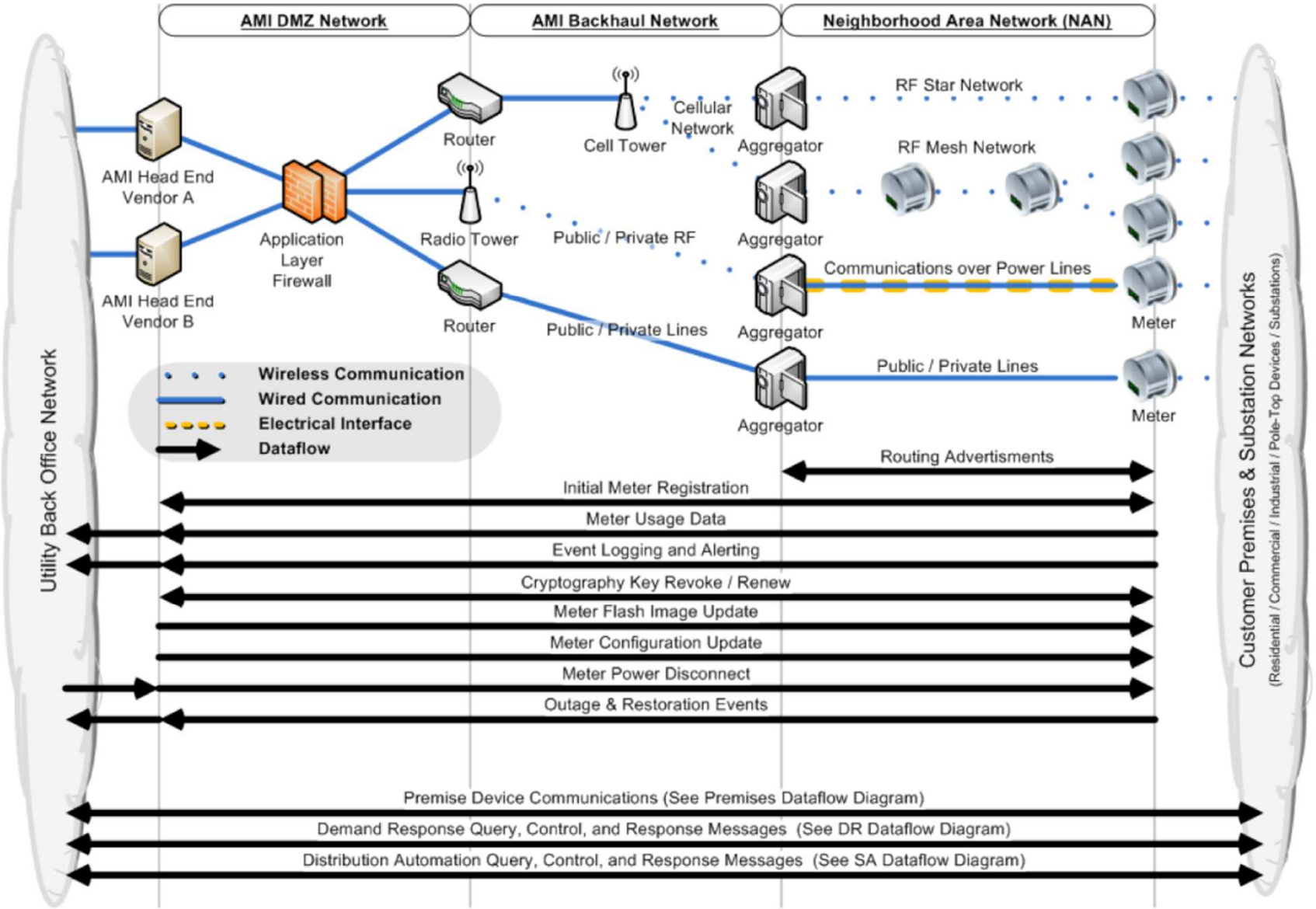
Jefferson Capovilla – CPqD
E-mail: jrodrigo@cpqd.com.br

Introduction

- CPqD: nonprofit R&D organization*
- Project: security analysis of smart meters
- Big data scenarios:
  - Smart grid dataflow
  - Binary firmware analysis

*CPqD:
http://www.cpqd.com.br/en/cpqd.html

Smart grid Dataflow



Smart grid dataflow

- Amount of data proportional to:
  - Number of smart meters installed in the grid
  - Control and monitoring messages
- Each communication channel presents attack points
  - Needs of security features for system reliability
- Data flow volume for 27 million smart meters*:

Meter Type	Single Meter p.a.	Total Meter Population p.a.
Electricity	Less than 1.5 MB	30 – 40 TB[36]

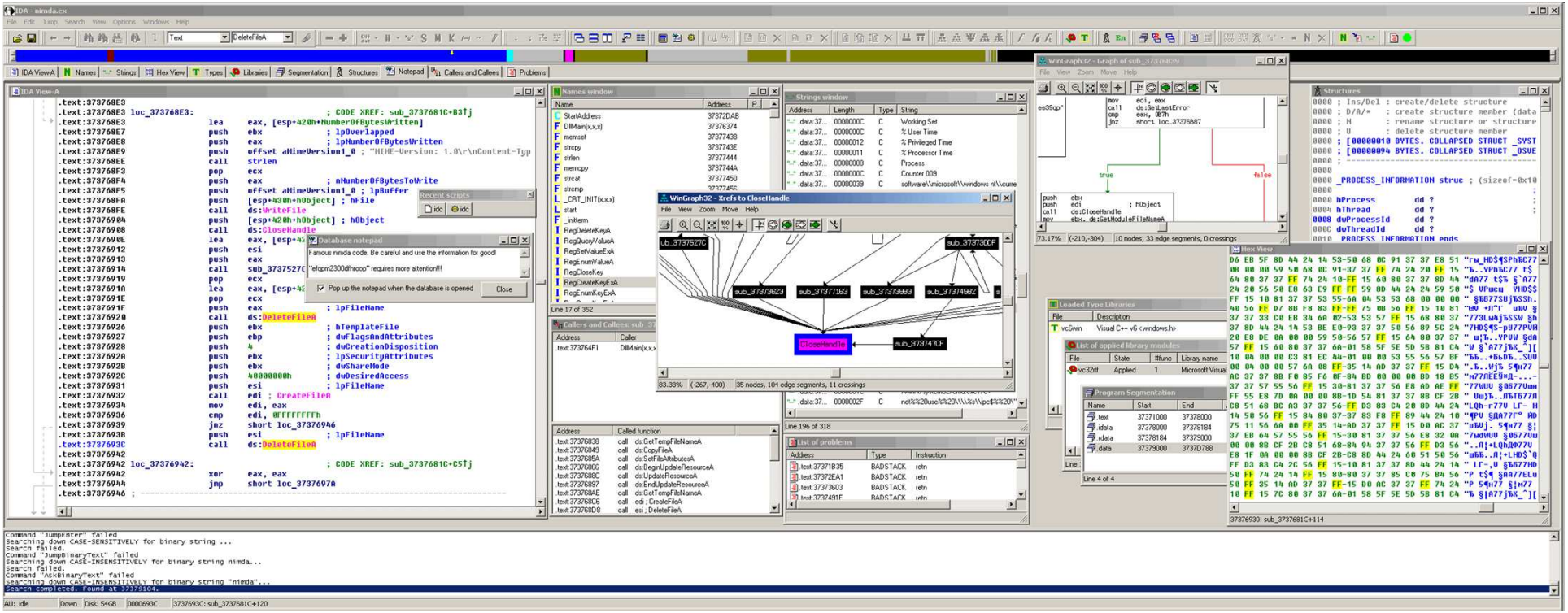
*Data extracted from: http://www.energynetworks.org/modx/assets/files/electricity/futures/smart_meters/ENA-CR008-001-1%204%20_Data%20Traffic%20Analysis_.pdf

Firmware binary analysis

- Firmware: embedded software running in smart meter
- As part of the methodology of smart meter security analysis:
  - Firmware binary extraction from non-volatile memory
  - Firmware binary disassembly(IDA-pro)
  - Firmware binary code analysis
  - Firmware binary exploitation
- Firmware assembly code: > 30000 lines
- Tends to increase with new functionalities in smart-meters
- Analysis performed manually with some tools help

Firmware Binary Analysis

- Example of firmware analysis using IDA-pro



Thank You!

Contact:
Jefferson Capovilla– CPqD
E-mail: jrodrigo@cpqd.com.br

*www.**cpqd**.com.br*

GRUPPO TELECOM ITALIA

Seville 2014

PANEL: Information, Information, Information… How Much can we Handle?

ICCGI 2014
June 22 - 26, 2014 -
Seville, Spain

TELECOM ITALIA

PANEL: Information, Information, Information… How Much can we Handle?

PANELISTS

- Arno LEIST, Massey University, New Zealand
- Félix BUENDÍA, Universidad Politécnica de Valencia, Spain
- Jefferson CAPOVILLA, CPqD, Brazil
- Antonio J. TALLÓN-BALLESTEROS, University of Seville, Spain

Structure of the Panel:
Introduction on the topic
5 minutes speech per panelist

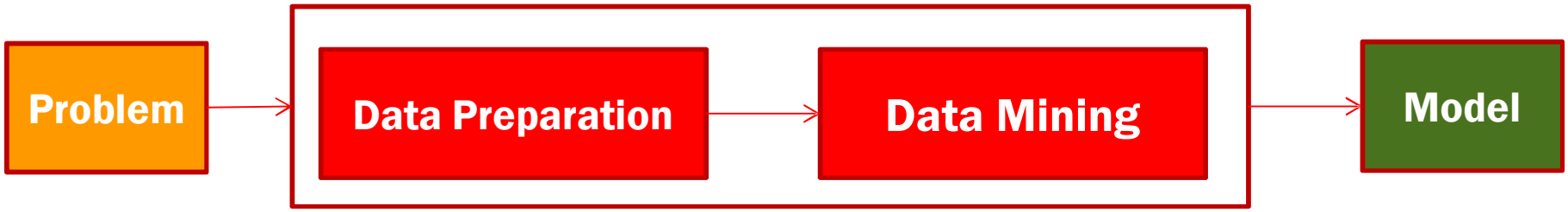
Discussion with the Audience is the key!

Information Processing in Machine Learning (IPML)

Antonio J. Tallón-Ballesteros, Ph. D.
Lecturer
Department of Languages and Computer Systems.
University of Seville (Spain)
atallon@us.es

Research Areas:
Data Mining, Supervised Learning, Neural Networks, Evolutionary Algorithms

IPML (1)



Knowledge Discovery in Databases (KDD)

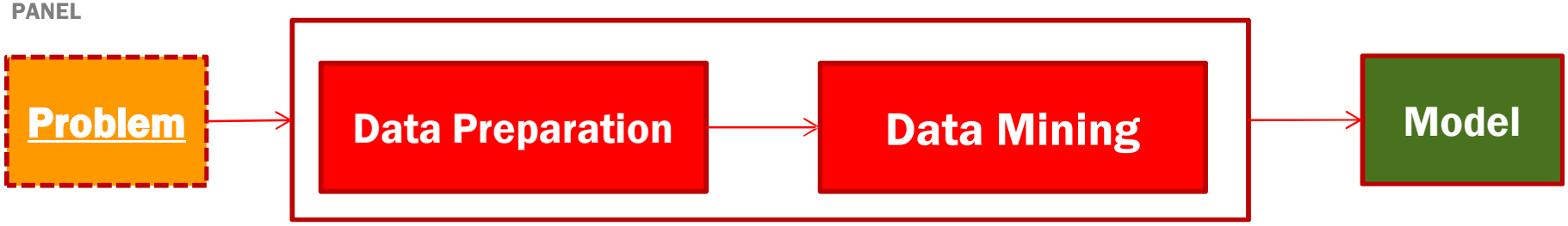
- **Complexity**
- ~~**Storage units** (gigas, ...)~~
- **Instances, features and/or classes**

- **Data cleansing** (missing, attribute noise*)
- **Feature Selection**
- **Instance Selection**

- **Specific tools: WEKA, Orange,...**
- **Classification approaches**
  - Decision trees (C4.5)
  - Based on Naive Bayes
  - Rule-based classifiers
  - k-Neareast Neighour (k-NN)

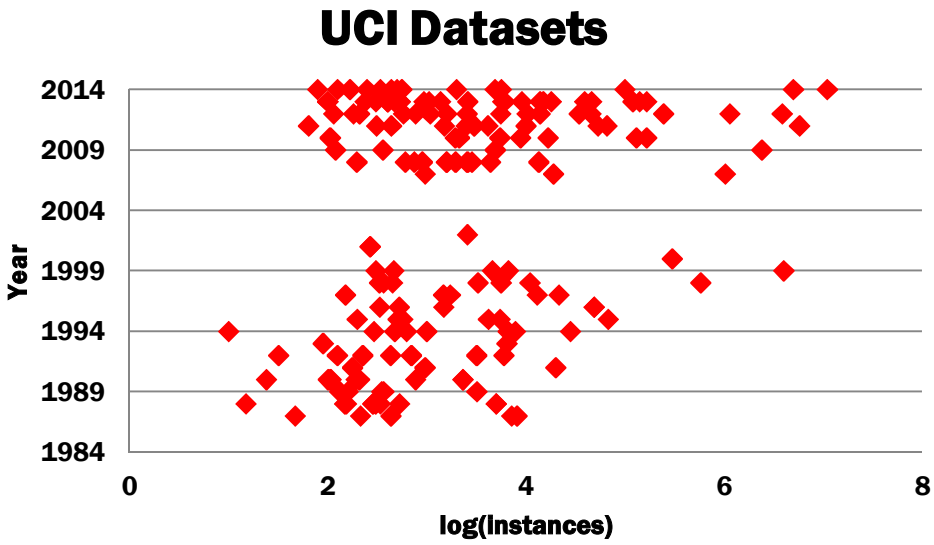
* Deleting or Keeping Outliers for Classifier Training? Tallón-Ballesteros, A.J. and Riquelme, J.C. (to appear in IEEE NaBIC 2014)

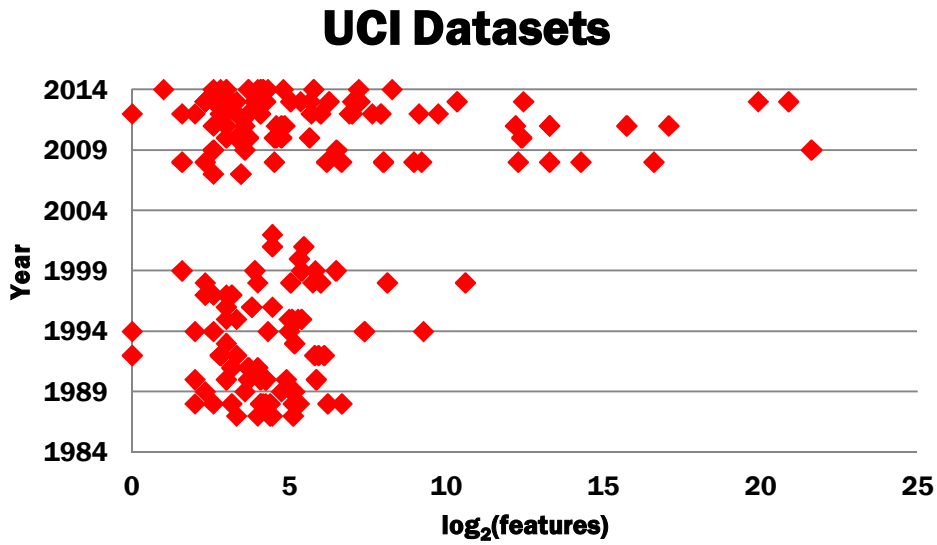
IPML (2)



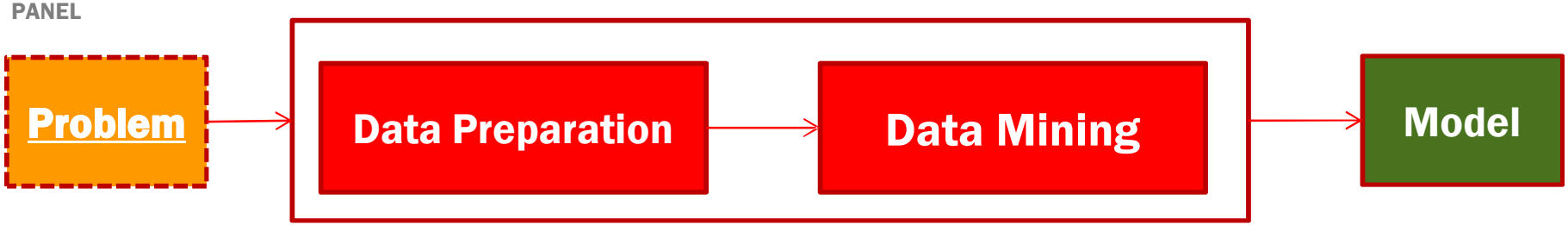
Knowledge Discovery in Databases (KDD)

- How much can we handle?
- UCI repository (Univ. of California at Irvine) → ~ 200 problems



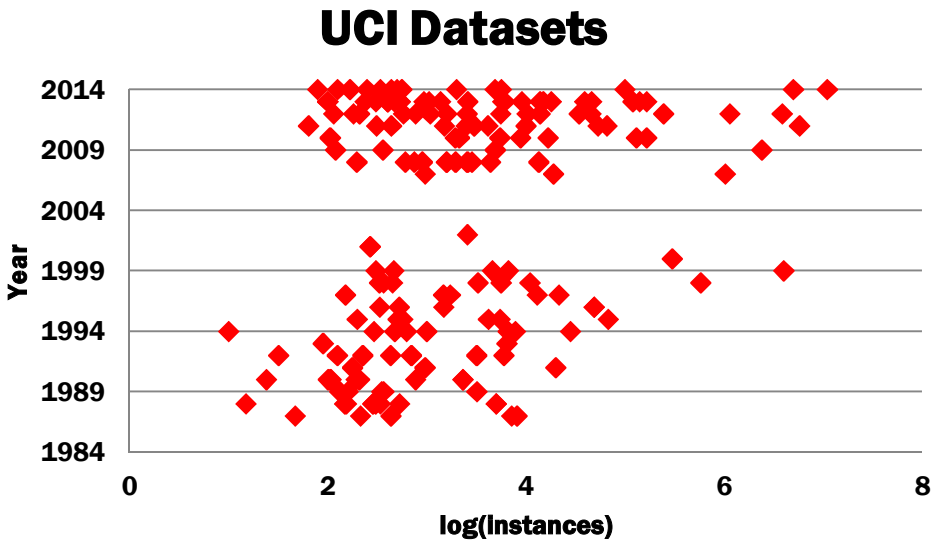


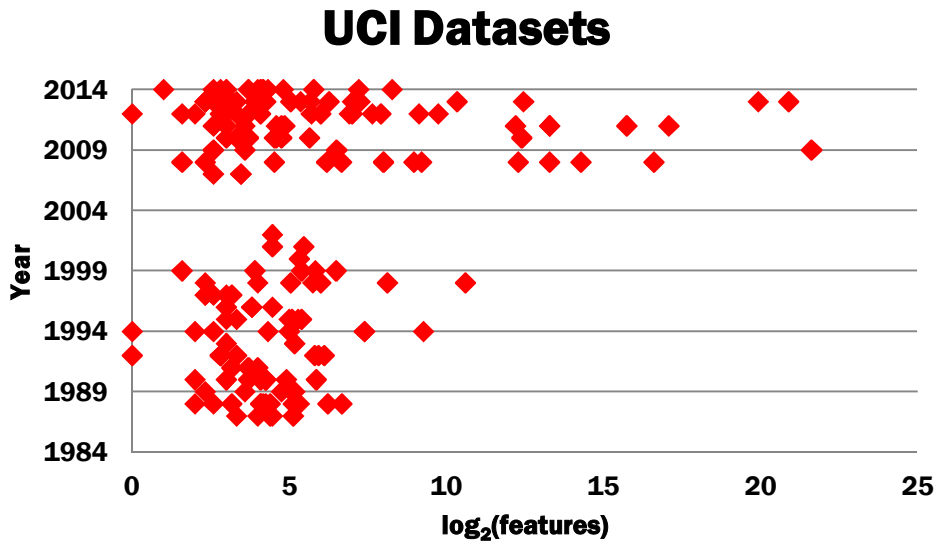
IPML (3)



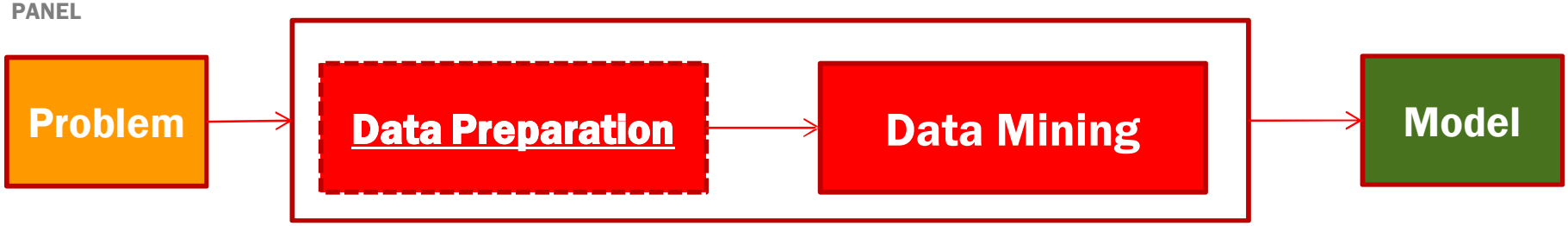
Knowledge Discovery in Databases (KDD)

- Number of instances is increasing (Top 1: 2014)
- Number of features stopped the increase 5 years ago (Top 1: 2009)





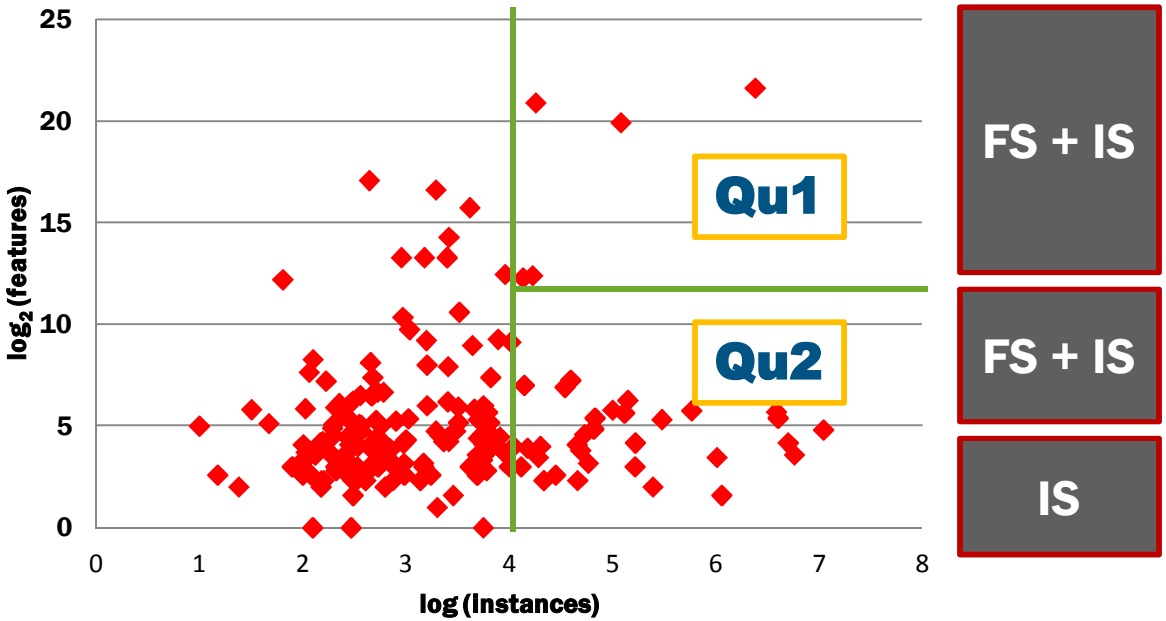
IPML (4)



Knowledge Discovery in Databases (KDD)

- **Which preparation techniques must be applied?**

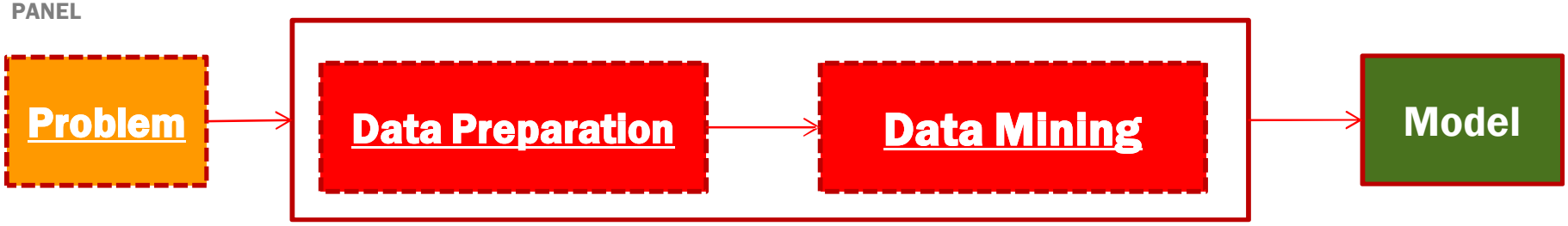
UCI Datasets



X=4 → ~ 10 Thousand INST.
X=6 → ~ 1 Million INST.

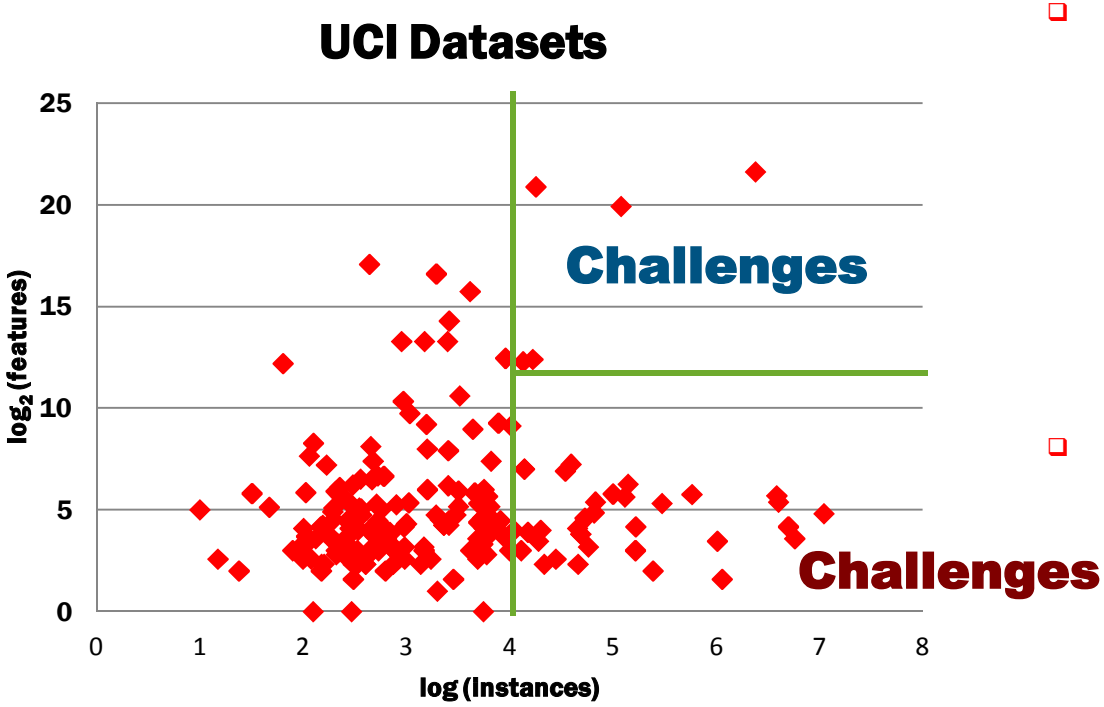
Y=5 → ~ 32 FEAT.
Y=10 → ~ 1 Thousand FEAT.

IPML (5)



Knowledge Discovery in Databases (KDD)

- **Top 3 problems**



- W.r.t. features:
  - URL Reputation (2009): Computer
  - Gas sensor (2013): Chemistry
  - YouTube (2013): Computer
- W.r.t. instances:
  - HIGGS (bossons) (2014): Physics
  - Record Linkage (2011): Computer
  - SUSY (particles) (2014): Physics