

# Data Mining for Transportation

Victor Ciriza, Lab

Manager

Advanced Development  
Lab

Xerox Innovation Group

## Xerox Overview

- FORTUNE 200 Company
- Leader in business process and document management solutions
- Over \$22 billion in annual revenue
- Over 130,000 employees
- Services business includes the

## Xerox Innovation Group

### Investment

- Xerox + Fuji Xerox = \$1.5B per year in R&D
- 55000 patents, >17800 active in USA
- ~10 patents per week
- One of the most profitable patents in history of US Pat Office

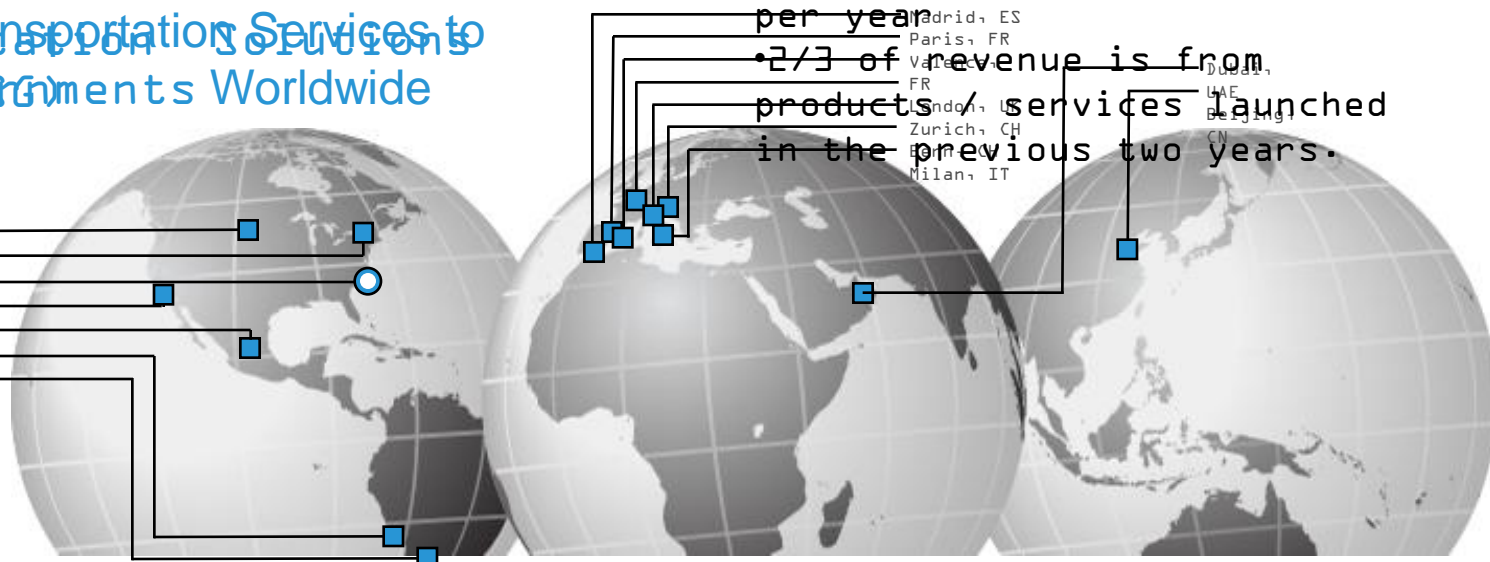
### Value of Innovation

- ~40 new products / services per year
- 2/3 of revenue is from products / services launched in the previous two years.

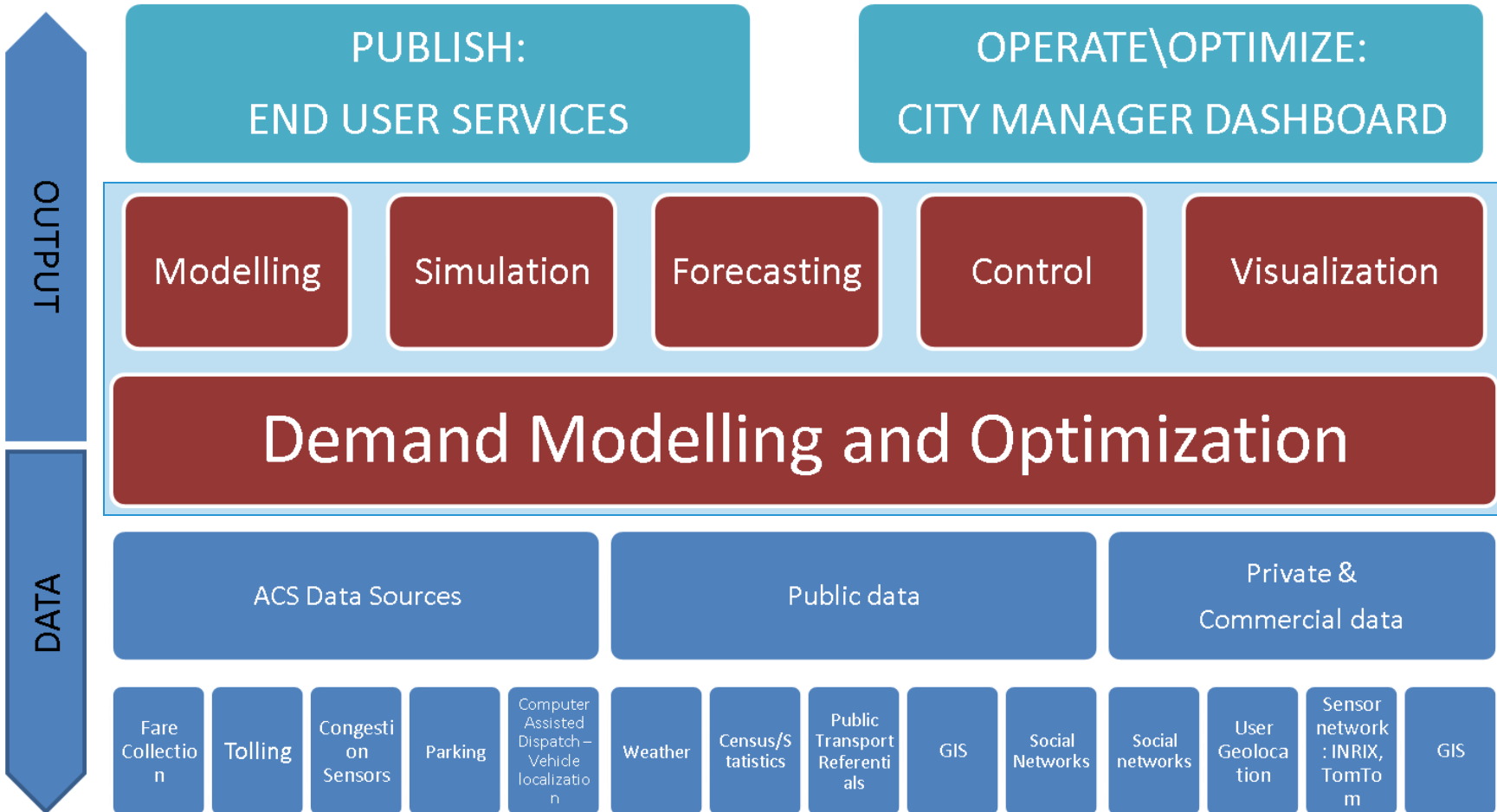
## #1 in Transportation Solutions to Group Governments Worldwide

### TSG Sales Offices

- Edmonton, CA
- Toronto, CA
- Washington DC, US
- Los Angeles, US
- Mexico City, MX
- Lima, Peru
- Santiago, CL
- Headquarters in Washington, DC metro area



# Data Mining – Multiple Sources and a Dual Goal



# Data Mining – Research & Engineering Challenges

## Operational

- **Multiple sources & Criticality:** Data warehousing
- **Privacy & Data Retention:** anonymization and policies
- **Volumes:** BI platforms, GPU and cluster computing

## Equipment & Ridership Analysis

- **Data Representation:** vector of features
- **Key factor detection:** Latent Semantic Analysis, LDA(Clusters)
- **Prediction:** regression models, semi-supervised models

## Validation Sequences Analysis

- **Robustness to missing data:** probabilistic generative models
- **Discrete choice:** probit models
- **Sequence dynamics:** Hidden Markov Models (HMM/FHMM)
- **Simulation:** MCMC, Gibbs sampling (eg:BUGS)

## Visualization

- **Geographical Context:** GIS based visualizations
- **Time Sequences:** heatmap animations on GIS
- **Dashboards:** Web 2.0 or RIA frameworks (eg: Flash)

# Demand Management – Automating Transportation Surveys



- **Cities as “organic entities”**
  - Urbanism changes
  - Population dynamics
  - Economic transformations



- **User behaviours & Surveys**
  - Required: qualitative needs
  - Expensive
  - Not quick

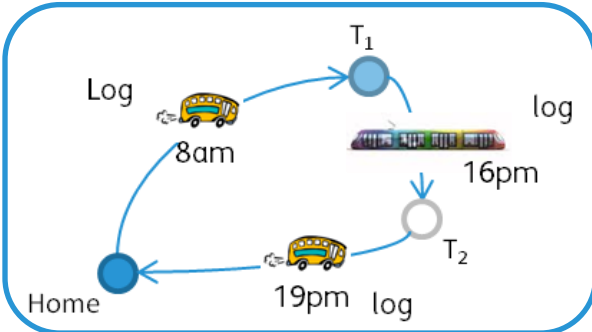


- **Opportunity: Contactless cards**
  - Card unique identifier: allow building travel profiles
  - Provide statistically significant samples
  - Near real time information
  - Automation

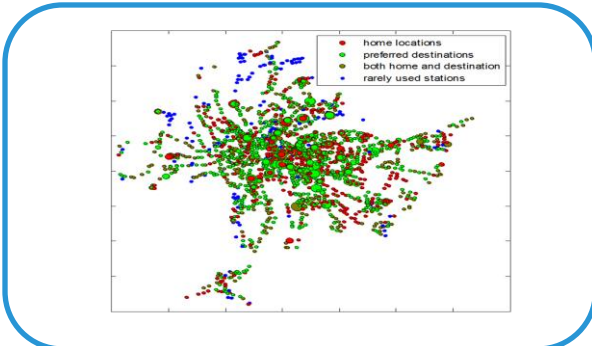
# Demand Management – Mining Fare Collection Data



- **Data Collection**
  - Contactless card transaction data warehouse
  - Minimum data: validator, user, timestamp
  - Seamlessly integrated in modern fare collection



- **Processing Transportation Usage Models**
  - Based on validation sequences
  - Model: reflects customer's ridership logic
  - Robustness: missing data, seasonality...
  - Probabilistic approaches fit best



- **City Dashboards**
  - Real-time visualizations of transport needs
  - Home-Work\Origin-Destination locations
  - Gaps and overcapacity identification
  - Simulations

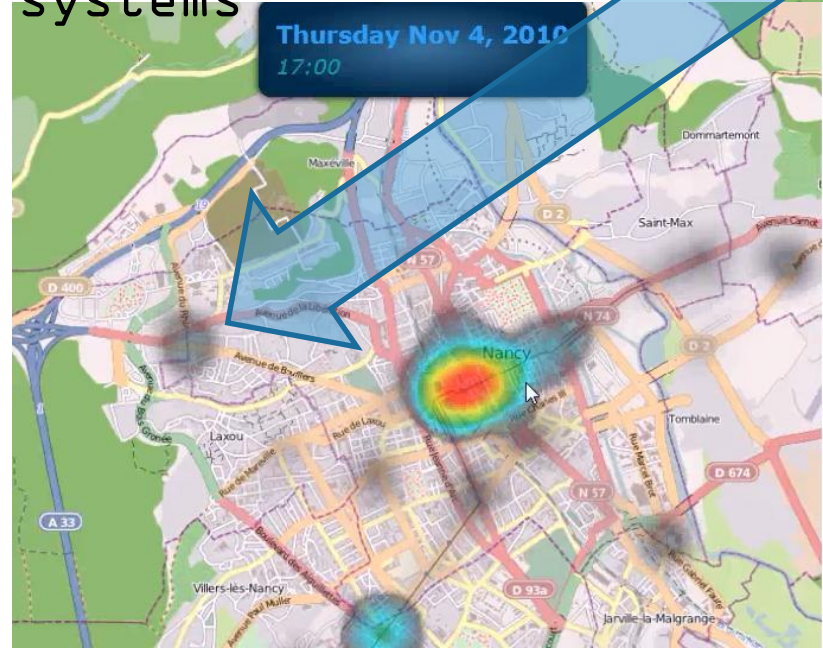




# Demand Management – ACS\Xerox City Dashboard

**Real example**

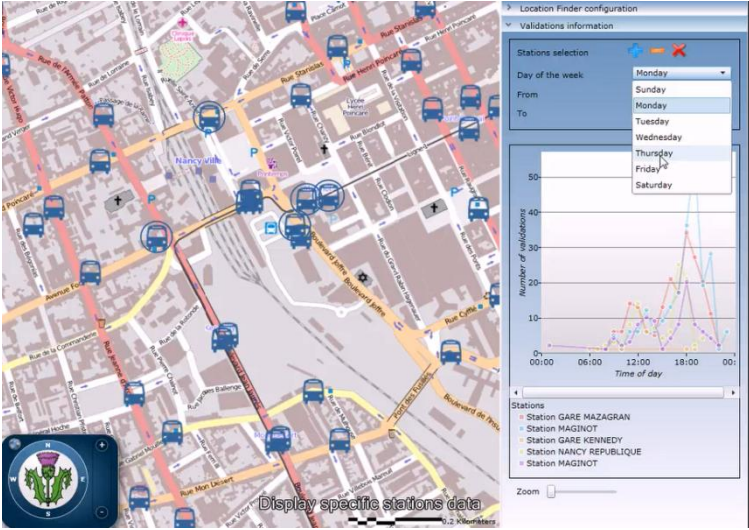
- French city
- Small\Medium network
- Multimodal
- ACS fare collection systems



**Visualizing**

- Overall validation activity
- Zoom-in for Home\Work profiling

# Demand Management - Conclusions



## Research

- Manual surveys validate fare collection data processing results
- Usage models progressing towards citywide dynamics simulation (eg: what if analysis)
- Equipment & Ridership Analysis

THANK YOU

## Development

- Customer Lead Innovation: close





**xerox**



# Introduction to Xerox R&D Capabilities

Research as an engine of growth      R&D Investments  
 1600 patents in 2010 (Xerox / Fuji Xerox) dedicated to R&D for Xerox  
 In the top 20 world leaders for patents Xerox

Xerox Research Centre of Canada  
 Mississauga, Ontario, Canada



Xerox Research Centre Europe  
 Grenoble, France



Palo Alto Research Center, Inc  
 Palo Alto, California, USA



Xerox Research Center  
 Webster  
 Webster, NY, USA



Xerox India Innovation Hub  
 Chennai, India



Fuji Xerox Japan



# Fare Collection

WORLDWIDE REFERENCES - Leverage footprint in territories



Melbourne Australia



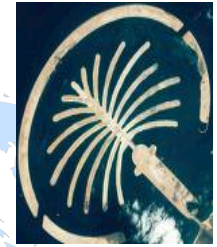
OeBB Austria



Vienna Austria



Calgary Canada



Dubai UAE



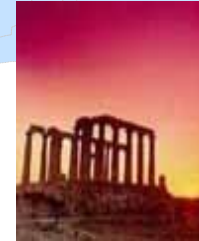
Hong Kong China



Berlin Germany



Mannheim Germany



Athens Greece



NS Holland



Kuala Lumpur Malaysia



NSB and Flytoget  
Norway



SBB Switzerland



Zurich Switzerland



New Jersey USA