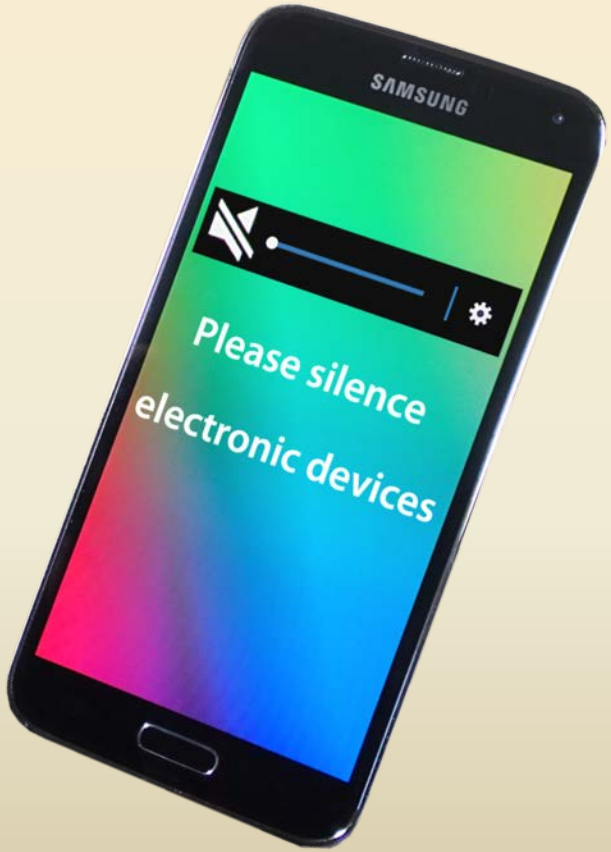
21st Annual GIS/CAMA Technologies Conference Chattanooga Convention Center

Continuing Education (CE) Credit

Recertification Credit forms for CE credit can be collected from the Registration Desk on <u>Thursday</u>.

Housekeeping

The conference proceedings will be available approximately 8 weeks after the conference.



Integrating Predictive Model Markup Language into a CAMA System

Experiences from the Maricopa County Assessor's Office

David Whiterell, RES
whiterelld@mail.maricopa.gov

Jennifer Rearich
rearichj001@mail.maricopa.gov

Overview

- Starting Line
  - Existing Processes
  - Background Information
- The Goal
  - Model Deployment using PMML
- From Concept to Actuality
  - Decision Process
  - Sowing Teamwork for Innovation
- Envisioning the Future
  - Alternative Models and PMML
- Code Examples

About Maricopa County

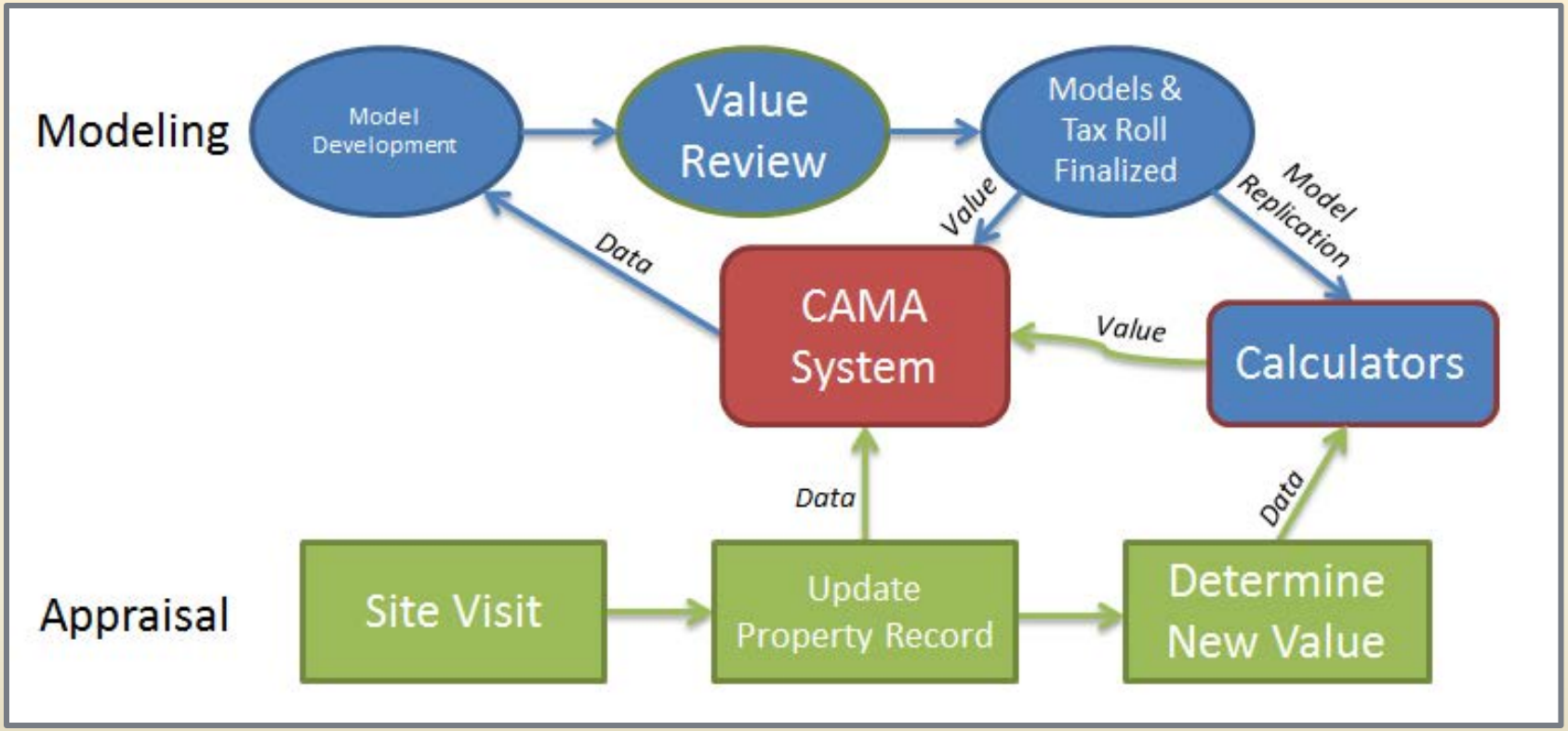
- 4.1 million residents (July 2015)
- 9,200 square miles
- 106.1°F avg summer temp
- 300 days of sunshine
- 1.58 million parcels
- $450 billion total FCV (2016)
- 260 Assessor's Office staff
  - 8 modelers
- Two Annual Assessment Rolls
  - Notice of Value (NOV)
  - Notice of Change (NOC)

Existing Processes

- Mid-1990s
  - Mass appraisal modeling program established
  - Explosive growth demanded new efficient processes for property valuations
- Statistical Package for the Social Sciences (SPSS)
- 71 annual multiple regression analysis (MRA) models produced in 2016

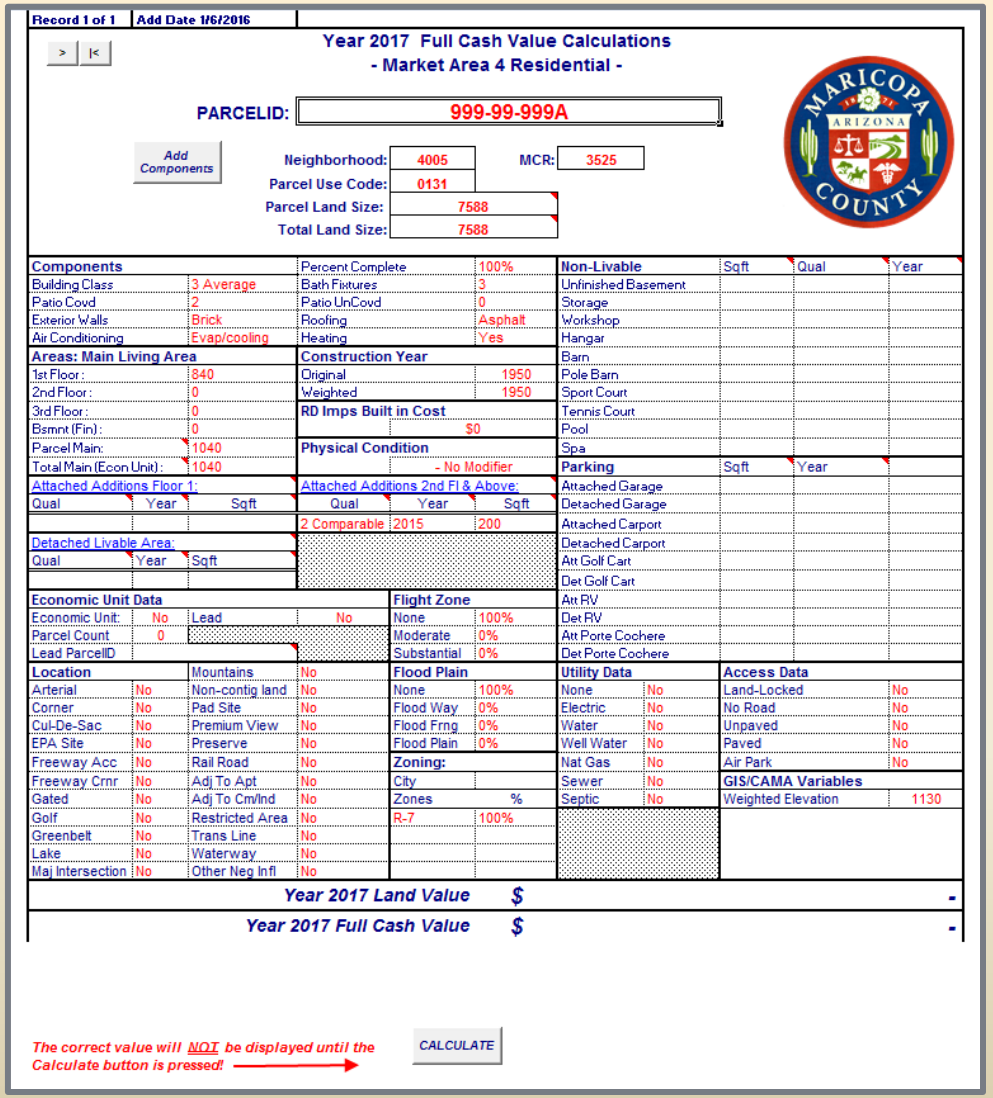
Current Model Types	
Residential	Condominium
Residential Land	Commercial Land
Apartment Income & Market	Industrial Market
Office Income & Market	

Existing Processes



Existing Processes

- Calculators
  - SPSS models replicated in Microsoft Excel
  - Allows appraisers to interact with models to update property values



Record 1 of 1 | Add Date 1/6/2016

> |<

Year 2017 Full Cash Value Calculations
- Market Area 4 Residential -

PARCELID: 999-99-999A

Add Components

Neighborhood: 4005 MCR: 3525
Parcel Use Code: 0131
Parcel Land Size: 7588
Total Land Size: 7588

Components		Percent Complete	100%
Building Class	3 Average	Bath Fixtures	3
Patio Covd	2	Patio UnCovd	0
Exterior Walls	Brick	Roofing	Asphalt
Air Conditioning	Evap/cooling	Heating	Yes

Areas: Main Living Area	
1st Floor:	840
2nd Floor:	0
3rd Floor:	0
Bsmnt (Fin):	0
Parcel Main:	1040
Total Main (Econ Unit):	1040

Construction Year	
Original	1950
Weighted	1950

RD Imps Built in Cost	
	$0

Physical Condition	
	- No Modifier

Attached Additions Floor 1:		
Qual	Year	Sqft

Attached Additions 2nd Fl & Above:		
Qual	Year	Sqft
2 Comparable	2015	200

Detached Livable Area:		
Qual	Year	Sqft

Economic Unit Data			
Economic Unit:	No	Lead	No
Parcel Count	0		
Lead ParcelID			

Flight Zone	
None	100%
Moderate	0%
Substantial	0%

Location		Mountains	
Arterial	No	Non-contig land	No
Corner	No	Pad Site	No
Cul-De-Sac	No	Premium View	No
EPA Site	No	Preserve	No
Freeway Acc	No	Rail Road	No
Freeway Crnr	No	Adj To Apt	No
Gated	No	Adj To Cm/Ind	No
Golf	No	Restricted Area	No
Greenbelt	No	Trans Line	No
Lake	No	Waterway	No
Maj Intersection	No	Other Neg Infl	No

Flood Plain	
None	100%
Flood Way	0%
Flood Frng	0%
Flood Plain	0%

Zoning:	
City	
Zones	%
R-7	100%

Utility Data		Access Data	
None	No	Land-Locked	No
Electric	No	No Road	No
Water	No	Unpaved	No
Well Water	No	Paved	No
Nat Gas	No	Air Park	No
Sewer	No	GIS/CAMA Variables	
Septic	No	Weighted Elevation	1130

Non-Livable	Sqft	Qual	Year
Unfinished Basement			
Storage			
Workshop			
Hangar			
Barn			
Pole Barn			
Sport Court			
Tennis Court			
Pool			
Spa			

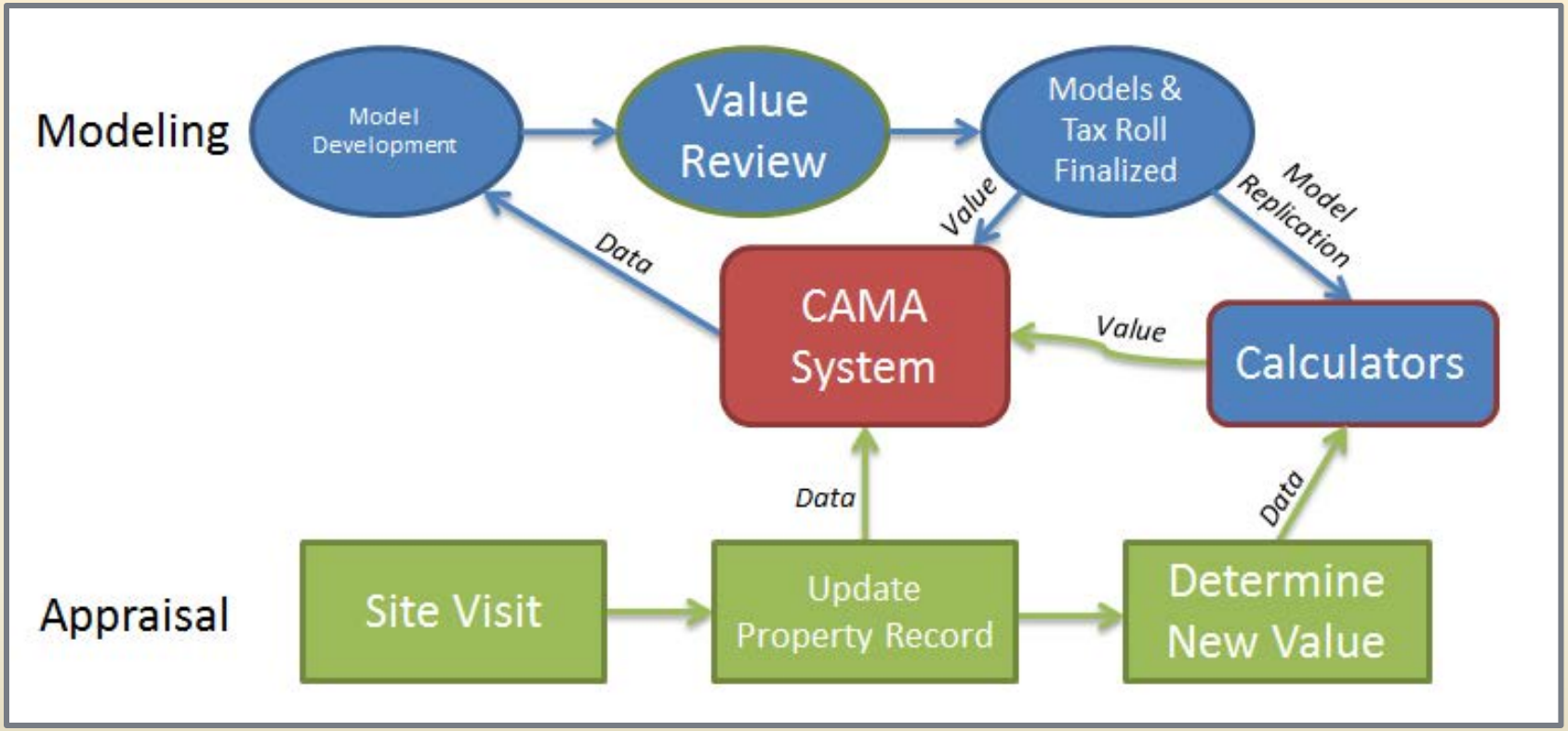
Parking	Sqft	Year
Attached Garage		
Detached Garage		
Attached Carport		
Detached Carport		
Att Golf Cart		
Det Golf Cart		
Att RV		
Det RV		
Att Porte Cochere		
Det Porte Cochere		

Year 2017 Land Value	$	-
Year 2017 Full Cash Value	$	-

The correct value will NOT be displayed until the Calculate button is pressed!

CALCULATE

Existing Processes



The Goal

- Integrate MRA models within CAMA System
  - Instant revaluation based on an assigned model
  - Creates more efficient appraisal workflow

- **Model Deployment**
  - Data Science terminology
  - Application of an existing model to new data
  - Many different model deployment methods available

IAAO Standard on AVMs

"An AVM must be tested to ensure that it meets required accuracy standards before being **deployed**."

(Section 2.3.6)

"The process of developing and **deploying** an automated valuation model must include safeguards to insure [sic] the accuracy of data used and the integrity of results produced."

(Section 8)

PMML Background

- Predictive Model Markup Language
- XML-based language
- Developed by Data Mining Group (DMG)
- Open standard for representing and sharing predictive models between different applications
- Version 0.7 developed in 1997; now on 4.3

PMML and Model Deployment

- Many industries (healthcare, insurance, banking, government) develop predictive models to determine credit risk, targeted audiences, fraud detection, quality control, etc

Making a Decision

- Considered numerous deployment methods
- Focused on three options:
  - Iron Python (Python integrated with .NET framework)
  - Custom vendor built program
  - PMML

Making a Decision

- Key questions:
  - Could the option successfully calculate within a CAMA System?
  - What impact would the option have on existing modeling processes?
  - How much time would it take to annually convert 70+ models from SPSS Syntax to PMML?
  - Could conversion be automated?
  - How quickly could values be calculated (both batch [anticipating upwards of 2 million parcels in the near future] and single)?
  - Was the option flexible in form and language (i.e., adaptive to non-regression model types or other statistical programs)?
  - Did the option present any potential database security issues?

Making a Decision

- PMML selected
  - Suitable for multiple types of predictive models
  - Compatible with numerous statistical programs
  - Portable for different operating systems and uses
  - Batch calculations surpassed speed expectations
    - Calculated 250,000 values in less than 2 minutes

Making a Decision

- Challenges of existing SPSS/PMML tools
  - SPSS Syntax allows certain types of unconventional coding
    - Overwriting original variables

      SQFT = SQFT – FINISHED_BASEMENT <-bad for PMML

      SQFTM = SQFT – FINISHED_BASEMENT <-good for PMML
    - SPSS Version 22 supports PMML 4.1
      - PMML 4.2 is preferable
      - Modeling processes could be adjusted to accommodate, but would require significant time and effort
  - Amount of work required to utilize existing SPSS to PMML conversion tools limited attraction

Making a Decision

- Positives of switching to R
  - Free, open source program for statistical computing
  - Supports PMML- conversion packages already exist
  - Increasingly taught in higher education
- Challenges of switching to R
  - Maintenance of SPSS-based legacy models
  - Modelers still need to complete some manual manipulation for PMML
    - Unless custom R package developed to support transformations and customizations required

Making a Decision

- Existing options presented problems
  - Modeling process dictated by statutorily-defined calendar
  - Not enough time to manually code PMML
  - Increasing staff size unlikely
- Envisioned ideal SPSS to PMML process
  - Long-term, stable, and easy to use solution
  - Modelers should remain modelers
    - Minimal time requirements for model deployment
  - Write PMML in parallel with modeling activities

PMML Methodology Decisions

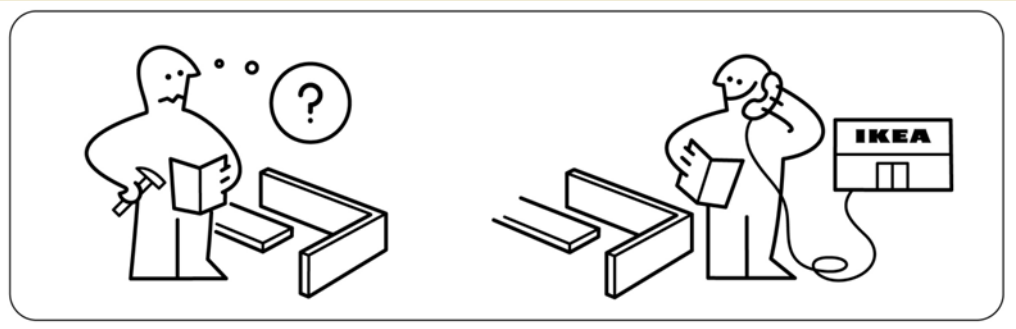
- Custom PMML Code Generator
  - Recognizes existing modeling processes
  - Eliminates teaching modelers how to manually manipulate PMML code
    - Reduces errors
    - Allows modelers to focus on modeling
- Verification Program
  - Ensures PMML calculated values = SPSS modeled values

PMML Methodology Decisions

- High-level understanding of database
- Supported model types
- Transformation mathematics
- Eight slightly different modeling styles
- Tests to ensure accuracy
  - Adheres to schema
  - Produces expected values

Learning the Fundamentals

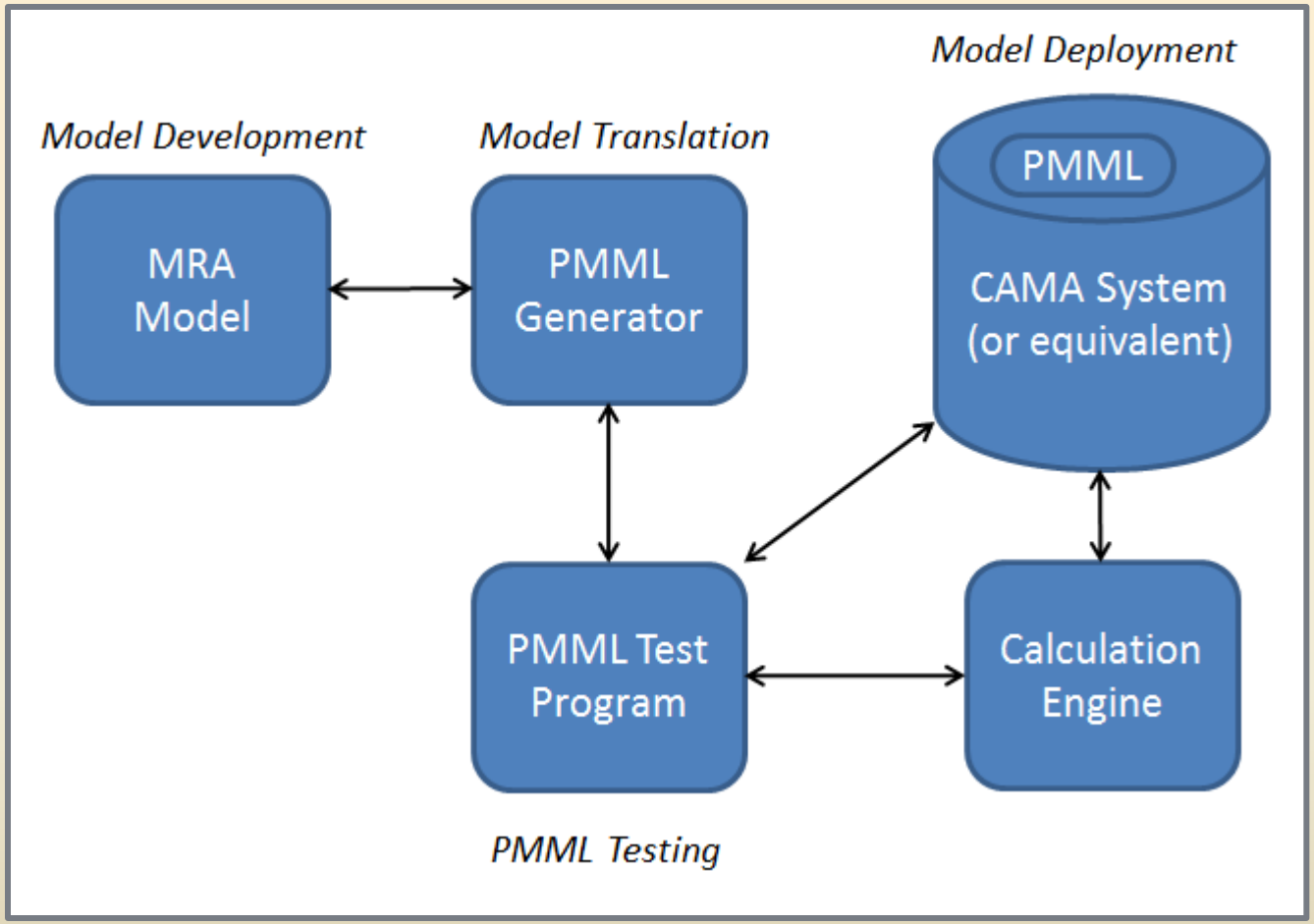
- Limited PMML resources
- PMML Class from UC San Diego Extension Program
  - *Predictive Models with PMML*
  - Conclusions:
    - Modelers are not computer programmers
    - Achieving end goal with PMML was going to be difficult



Learning the Fundamentals

- Basic programming training for modelers
- University of Michigan + Coursera MOOC
  - Series of online Python Courses
  - Collectively applied knowledge to familiar scenario
- Established foundational programming skills necessary to read code and troubleshoot

Transforming Processes



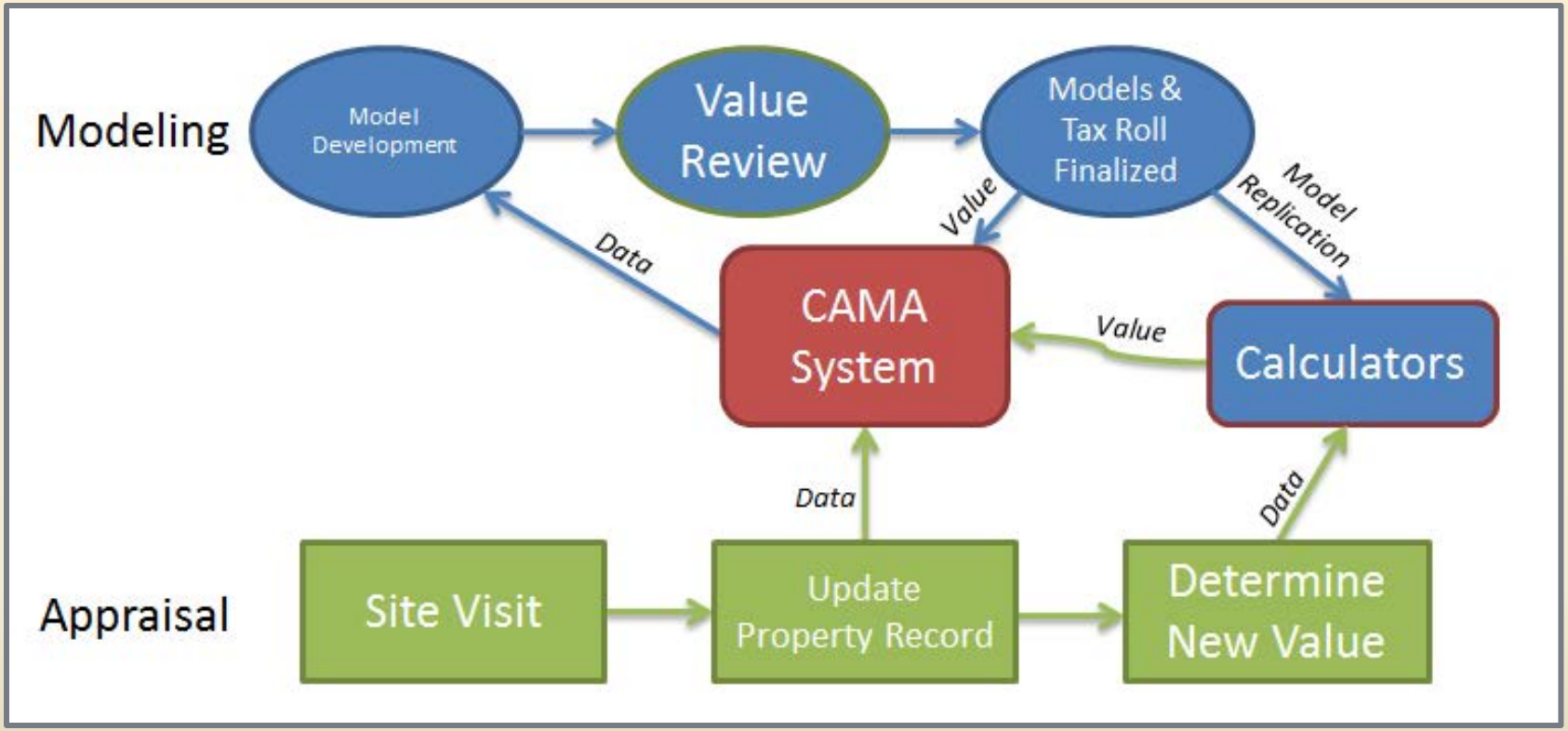
Innovation has nothing to do with how many R&D dollars you have. When Apple came up with the Mac, IBM was spending at least 100 times more on R&D. It's not about money. It's about the people you have, how you're led, and how much you get it.

- Steve Jobs, *Fortune,* November 9, 1998

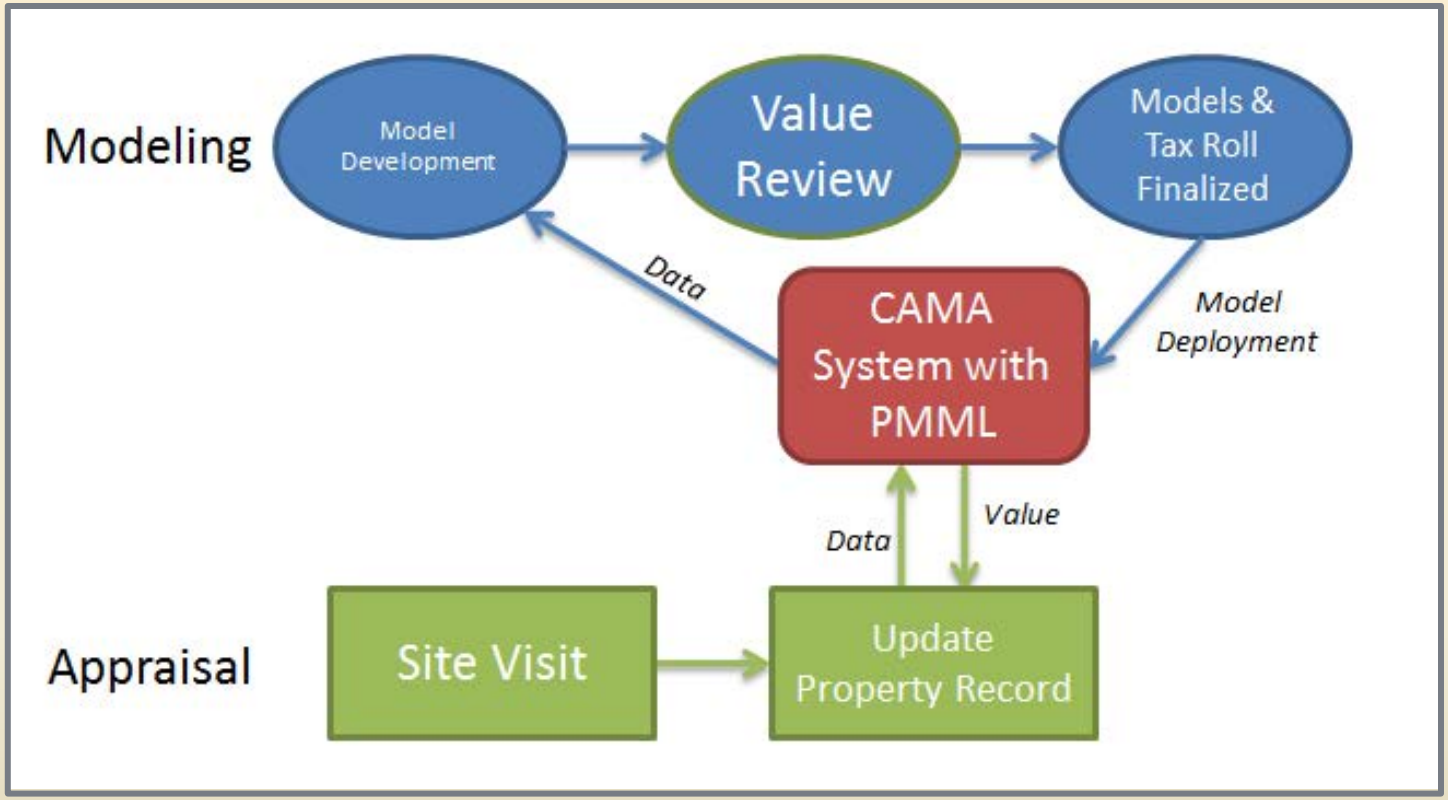
Sowing Teamwork for Innovation

- Workload assignment
  - PMML Project added to overall workload
  - Each modeler took on additional work
- Change and challenges
  - This wasn't easy
  - Many roadblocks along the way
    - Lots of discussion and spirited debate
- The Point
  - Keep focused on the goal and future benefits

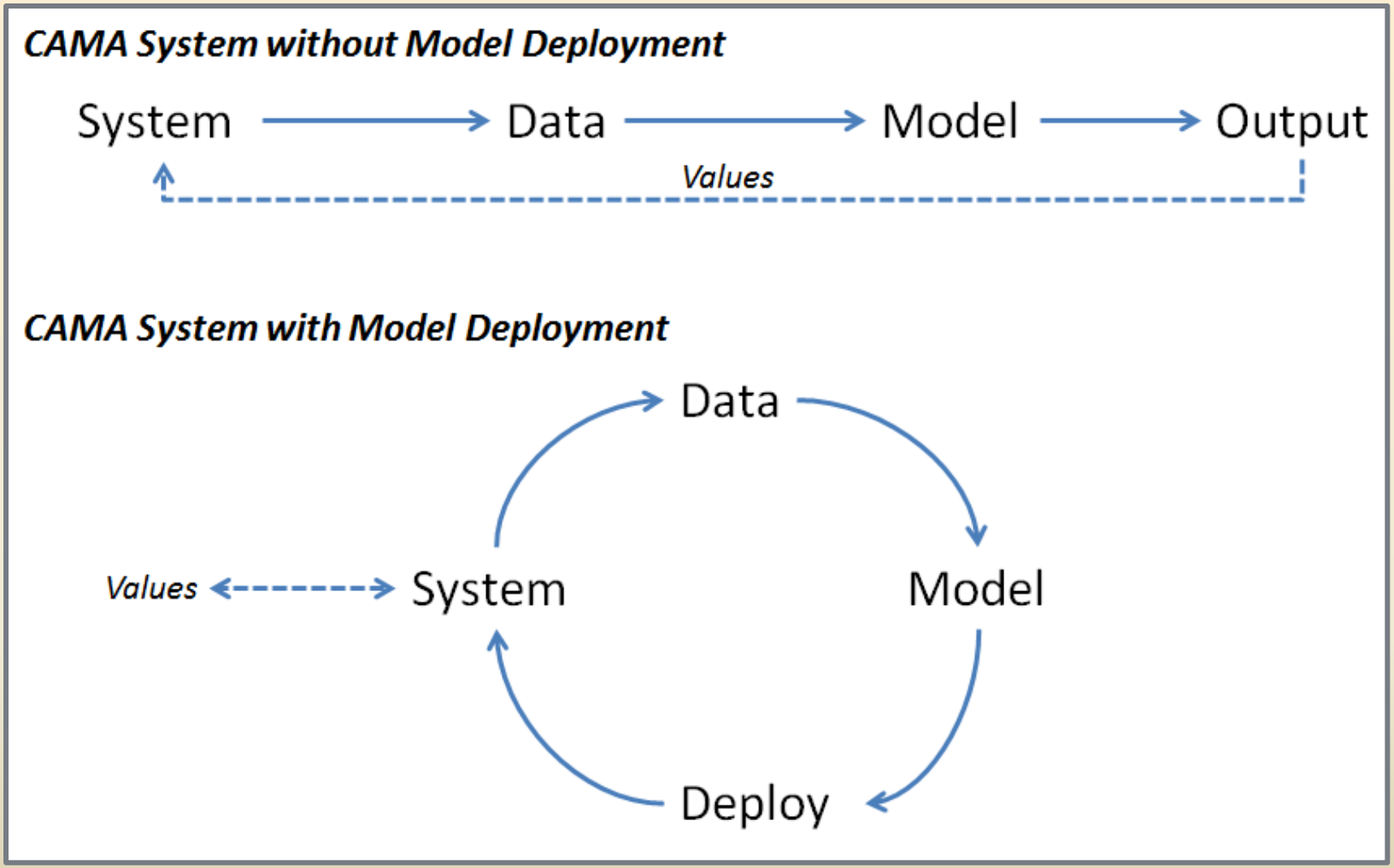
Existing Processes



Envisioning the Future



Envisioning the Future



Envisioning the Future

- There are all types of predictive model types supported by PMML
  - Deploying PMML-based predictive models for other mass appraisal uses
    - Comparable Sales
    - Data Collection- Effective Age scenarios
    - Data Integrity- Sales Verification
    - Decision Processes- Routing permits or appeals

Example of a Recode Statement

SPSS Syntax:	**R Code:**
recode nbhd (15006=1) (else=0) into nbhd15006.	data$nbhd15006 <- ifelse(data$nbhd== 15006,1,0) **where data is the name of the data frame*

PMML Code:

```
<DerivedField dataType="double" name="nbhd15006" optype="continuous">
    <Apply function="if">
    <Apply function="equal">
        <FieldRef field="nbhd"/>
        <Constant dataType="double">15006</Constant>
    </Apply>
        <Constant dataType="double">1</Constant>
        <Constant dataType="double">0</Constant>
    </Apply>
</DerivedField>
```

Example of a Compute Statement for a Continuous Variable

SPSS Syntax:	**R Code:**
compute sqftm = sqft – finished_basement.	data$sqftm <- data$sqft – data$finished_basement **where data is the name of the data frame*

PMML Code:

```
<DerivedField dataType="double" name="sqftm" optype="continuous">
    <Apply function="-">
        <FieldRef field="sqft"/>
        <FieldRef field="finished_basement"/>
    </Apply>
</DerivedField>
```

Example of a Compute/If Statement for a Binary Variable

SPSS Syntax:

```
compute lake_n4 = 0.
if (nbhd eq 10004 and lake eq 1) lake_n4 = 1.
```

R Code:

```
data$lake_n4 <- ifelse(data$nbhd==10004 & data$lake==1, 1 ,0)
```

**where data is the name of the data frame*

PMML Code:

```
<DerivedField dataType="double" name="lake_n4" optype="continuous">
   <Apply function="if">
      <Apply function="and">
         <Apply function="equal">
            <FieldRef field="nbhd"/>
            <Constant dataType="double">10004</Constant>
         </Apply>
         <Apply function="equal">
            <FieldRef field="lake"/>
            <Constant dataType="double">1</Constant>
         </Apply>
      </Apply>
      <Constant dataType="double">1</Constant>
      <Constant dataType="double">0</Constant>
   </Apply>
</DerivedField>
```

Key Points

- Model Deployment is now achievable
  - Mass appraisal models and CAMA Systems can now be integrated
  - Improved work processes benefit both modelers and appraisers
  - PMML makes models portable
- New potential uses for predictive models in mass appraisal environment

Questions?

Look for us:

- In Fair & Equitable
- At the 2017 IAAO Annual Conference in Las Vegas

Acknowledgments:

The Honorable Paul Petersen, Maricopa County Assessor

Tim Boncoskey, Chief Deputy Assessor

Eric Bails, Chief Technology Officer

Uwe Hohoff, Chief Modeler

Doug Pack, Modeler + Lead Programmer

Lance Hull, Brad Patton, John Mulvihill, Tyson Dziob, Modelers

References

Guazzelli, A., K. Stathatos, and M. Zeller. 2009. Efficient Deployment of Predictive Analytics through Open Standards and Cloud Computing. *SIGKDD Explorations* 11(1) 32-38. http://www.kdd.org/exploration_files/p5V11n1.pdf (accessed Oct. 21, 2016).

Guazzelli, A., M. Zeller, W. Lin, and G. Williams. 2009. PMML: An Open Standard for Sharing Models. *The R Journal* 1(1) 60-65. https://journal.r-project.org/archive/2009-1/RJournal_2009-1_Guazzelli+et+al.pdf (accessed Oct. 21, 2016).

Guazzelli, A., W. Lin, and T. Jena. 2012. *PMML in Action: Unleashing the Power of Open Standards for Data Mining and Predictive Analytics*, 2nd ed. CreateSpace Independent Publishing.

IAAO. 2003. *Standard on Automated Valuation Models (AVMs)*. Kansas City, MO: International Association of Assessing Officers.

Jena, T., A. Guazzelli, W. Lin, and M. Zeller. 2013. The R pmmlTransformations Package. Proceedings of the KDD 2013 PMML Workshop, August 2013, Chicago, IL. https://kdd13pmml.files.wordpress.com/2013/07/jena_et_al.pdf (accessed Oct. 21, 2016).

Pechter, R. 2009. What's PMML and What's New in PMML 4.0?. *SIGKDD Explorations* 11(1) 19-25. http://www.kdd.org/exploration_files/p3V11n1.pdf (accessed Oct. 21, 2016).

Severance, C. University of Michigan. *Python for Everybody Specialization* [Coursera MOOC]. https://www.coursera.org/specializations/python (accessed Jan. 5, 2017).

University of California, San Diego Extension. *Predictive Models with PMML* [CSE-41184]. http://extension.ucsd.edu/studyarea/index.cfm?vAction=singleCourse&vCourse=CSE-41184 (accessed Jan. 5, 2017).

U.S. Census Bureau. 2015. State & County Quickfacts: Maricopa County, AZ. http://www.census.gov/quickfacts/table/PST045215/04013 (accessed Jan. 17, 2017).

Wherli, J. 2016. *Predictive Model Markup Language (PMML): Unifying CAMA, AVM, GIS, & 3rd party analytics*. Proceedings of the 20th Annual GIS/CAMA Technologies Conference, Feb. 2016, Savannah, GA.

Whiterell, D. 2011. *Visualizing Assessment Data using low Cost Solutions*. Proceedings of the 2011 IAAO International Conference on Assessment Administration, Sep. 2011, Phoenix, AZ.


GIS-Pro 2017
October 23-26, 2017
Jacksonville, Florida


Assessment Leadership
Beyond All Limits
IAAO LAS VEGAS
ANNUAL CONFERENCE
& Exhibition
September 24-27
2017
MGM