An Activity-Based Semantic Urban Model Toolkit

Bob Schloss, Biplav Srivastava, Rosario Uceda-Sosa, Randy Kubich
IBM Research

2012 Apr 19, 1000CT
Assembling the Smarter City ... today’s cities

The data/information, services, structures and activities in cities are described in different ways by the agencies which manage them, by the denizens which use them and by companies which provide them.

This cacophony of views and understanding drive many of the inefficiencies, create uneconomic costs in cities and can limit the pace of growth in a city’s quality-of-life.

This work provides structures for the data, and the city infrastructure and services which reduce this confusion, providing applications which we expect will significantly reduce costs and improve life in the city.
There are significant associated research challenges

Open access to data requires automated, scalable approaches to data release and management (starting with security & privacy)

A sustainable ecosystem requires robust and scalable tools and interfaces to support interaction with and use of data and services

Unlocking value requires scalable semantic technologies to identify and organise contextual linkages across and within cross domain data
What do I mean by “a model”

Some information, supported by some software, used for 1 or more purposes
Some of the information partially represents entities in the cities
Models are approximations
Some of the information partially represents measures of city goals
Can either represent the city at a “moment in time” OR
Can represent one or more trajectories of the city through time
Does not have to represent the current real city – useful for alternatives

There is an ART to deciding what entities, influences to include in a given model
Too much detail is not helpful; too little detail can be problematic
There is a discipline to combine separately constructed models –
  Composition
  Coupling
But neither Composition nor Coupling can happen without a few connection points
Until now, 2D space and time have been the ground connection points
(For example, take a look at map-based mashups on the Web)
This should support simulation/predictive models, but unifying those efforts is an effort being recommended by OECD Global Science Forum and others

[Urban Systems modelling as a discipline] is hampered by lack of systematic knowledge exchange internationally (between model makers and users) and lack of system-wide independent review of both model-building methodologies and practice applications. A response is required to redress this imbalance as a matter of urgency. .....

Urban systems modeling has the potential for international recognition among cities and urban regions as a discrete and specialist field, as is the case with the well-established field of Building Information Modelling… The field of urban systems modelling could become known, for example as “City Information Modelling”, but first requires internationally recognised expert capacity, organisational mechanisms and cross-disciplinary standards of practice. At the technical level, tools such as CityGML, among others, might form the basis for such developments; as a commonly used existing information model, it can represent cities in 3D at various levels of detail…
Most models have some strategy to present, visualize – but I won’t talk about this.

Urban population (2009): 1.4 million
Urban population density: 2,150 persons/km
Real GDP\(^1\) (2009): USD7.2 billion
Real GDP annual growth (2005-09): 13.2%
Average real income per capita: USD1,998

However, models which are augmented by “standing queries” to NOTIFY people or organizations when certain thresholds have been crossed, or will be crossed are of particular interest.
We at the beginning of our a journey of sharing our data.

A lot of this data created or captured by the property owner.

This data (outside the boundary) created by a neighborhood group, or the city.
Goal of ISMP

- Develop Modular, Interconnectable Composable Information+ Models to be used (by IBM Researchers and Academics *initially*) in building Intelligent Cities applications, basing them on minimal modifications of (refactorings of) open de-jure or de-facto standards
  - Anticipate the Scenarios where information collected by one vendor, utility, agency or community-network is shared in near-real-time with other vendors, utilities, agencies ... even though the actual producing or consuming software may be heterogeneous
  - Broaden the conversation among the IT community from single-purpose data exchange protocols in the direction of grounded Linked Data
Scope of ISMP

- Provide software modelers a comprehensive language covering the 7 essentials (Represent-Connect-Notify-Query-Project-Produce-Consume-Measure)
- covering actions and roles of government agencies, businesses, individuals and their assets in space and time
- And which is consumable (not that hard to learn), coherent, comprehensive
An analogy
IBM Research has a deep desire to partner, validate, then standardize.

IBM produces:

- ISMP Infrastructure (2011 & 1H2012)
- ISMP for Academics w/simple transport (8/30/2012)
- ISMP and IIC for Academics & Non-IT (no charge) (10/10/2012)
- Digital Dublin or Digital Chicago or Digital Delhi Or Digital NYC (6/2013)
- I.S.M.P. in IOC, IWM, IPS (9/2013) Subject to change

The End

2015

ISMP Core
Donated to OASIS, W3C o New Open-City-Data- Consortium (1/2014)
### 10 Key Aspects of any Model Toolkit

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>How is it represented?</td>
<td>Intelligent Semantic Model Palette</td>
</tr>
<tr>
<td>Connections/context implicit or explicit?</td>
<td>OWL (Web Ontology Language)</td>
</tr>
<tr>
<td>Explicit [may invoke inferencing to guarantee all connections]</td>
<td></td>
</tr>
<tr>
<td>How updated/read by software?</td>
<td>JENA API, SPARQL</td>
</tr>
<tr>
<td>How is it persisted?</td>
<td>Graph Store or Relational databases</td>
</tr>
<tr>
<td>How is an instance exchanged?</td>
<td>N3 or Turtle (Linked Data)</td>
</tr>
<tr>
<td>How is it visualized?</td>
<td>[Not part of our Research]</td>
</tr>
<tr>
<td>How much detail?</td>
<td>Essential information about all spaces, buildings, throughfares, assets, resources, roles, natural conditions, incidents, service requests, measurements</td>
</tr>
<tr>
<td>How is variation from 1 city to another represented?</td>
<td>Customization, Extension packages</td>
</tr>
<tr>
<td>Connected to standards?</td>
<td>Yes: CAP, NIEM, NOAA...</td>
</tr>
<tr>
<td>Learnability?</td>
<td>Profusely documented, example data, example queries, tutorials</td>
</tr>
</tbody>
</table>

**Open World**
## 10 Key Aspects of any Model Toolkit

<table>
<thead>
<tr>
<th>Aspect</th>
<th>CityGML goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>How is it represented?</td>
<td>GML</td>
</tr>
<tr>
<td>Connections/context implicit or explicit?</td>
<td>?</td>
</tr>
<tr>
<td>How updated/read by software?</td>
<td>?</td>
</tr>
<tr>
<td>How is it persisted?</td>
<td>XML ?</td>
</tr>
<tr>
<td>How is an instance exchanged?</td>
<td>XML</td>
</tr>
<tr>
<td>How is it visualized?</td>
<td>?</td>
</tr>
<tr>
<td>How much detail?</td>
<td>?</td>
</tr>
<tr>
<td>How is variation from 1 city to another represented?</td>
<td>?</td>
</tr>
<tr>
<td>Connected to standards?</td>
<td>Yes: GML</td>
</tr>
<tr>
<td>Learnability?</td>
<td>?</td>
</tr>
</tbody>
</table>
The parts of the I.S.M.P. Ecosystem

1. The model by itself provides a smart dictionary of the common vocabulary (what types of events do we have, what kinds of stakeholders have competencies in public safety). Definitions appear in InfoSphere Business Glossary. Extended with domain models produced by others in step 6 in 2013 and beyond.
2. With instance access, questions like ‘what events in the city of Chicago are affected by road work in Main St on 7/15?’ can be answered - without having to permanently transfer data from existing data sources to the model.
3. Navigation tools allow any user (including apps) to query the model and get a non-ontology result or a filtered view.
4. Natural language definitions of data from other agencies is available to query construction and result examination time
5. Standard authoring tools can be used for ontology customization and management. New tools will be produced for collaborative evolution and for model-driven generation
6. Other teams in Academia, or among Professional Service Firms, build domain models which extend ISMP Core.
One **value** of the I.S.M.P. Ecosystem

One goal of a data model should ultimately be the cheap (2), standard (1,4) integration of data sources, providing a simplified access to them (3)

1. **ISMP Data Model (concepts)**
   - Standardization of vocabulary and app logic (needs to be synched with data)
   - Facilitates a robust implementation (helps detect inconsistencies)

2. **Data (Instance) access**
   (a) Cheap/fast data integration
   (b) No disruption to existing data
   (c) Add an additional source of the same kind of data, revising no queries (stretch)

3. **Navigation and query tools**
   (a) Standardization of data output
   (b) Isolation of ontologies for general users

4. **Glossary**

5a. **Authoring tools – std**
    Standard OWL and XML tools (Rational Design Mgr, TopBraid, Protégé, Pellet, SPARQL, etc.)

5b. **Authoring tools - new**
    Collaborative evolution tools
    Model-driven generation tools
    Customize-to-City tools
The deliverables of the ISMP Ecosystem and their timing

At different levels:
1. **Modeling**
   - CORE v2: end of April, Thoroughfares: end of August
2. **Modeling maintenance/tools**
   - November 2012
3. **Instance access runtime**
   - IIC 4Q2012
4. **Query/Navigation tooling**
   - 2013 (probably)

**1. ISMP Data Model (concepts)**
- Core (April 22, 2012)
- Thoroughfares, Characteristics of People
- Processes, Resource-pools, Svc Res’vtns
- Sensors
- Transport

**2. Data (Instance) access: MIDO**
- a. Prototype with research assets Nov, and Dec (Read-Only, Scalability Low, Rela’tml)
- b. Intersection with IIC end-of-the-year
- c. Might change query interface used by Eoin, Magda, Joey, Business Partners
- d. Write-version, Scalability High 2013 (?)

**End User Report Developer**

**Application Developer/Consultant**

**3. Navigation and query tools**
- TBD probably 2013

**4. Glossary**

**6. ISMP Domain Data**
- From other USC colleagues

**5a. Authoring tools – std**
- Standard OWL and XML tools
  - Rational Design Mgr, TopBraid, Protégé, Pellet, SPARQL, etc.

**5b. Authoring tools - new**
- Collaborative evolution tools
- Model-driven generation tools 4Q12
  - (e.g. Owl-enabled InfoSphere Data Architect)
- Customize-to-City tools 2013

**0b. Structured Scenarios (1Q12 & Only)**

**(Existing) City Data Sources**

- Database
- XML
Principal Use Cases - for City & Citizens
Validating the Open Innovation Platform using a Dublin collaboration

- Develop and deploy new products & services in a socially synergistic ecosystem
- Make more informed decisions about resources & investments
- Make current activities & operations more efficient & adaptive to continuously evolving conditions

Develop differentiators for our city, leveraging all credible stakeholders
Planning (Workforce, Capital Expenses, Supplies, Landuse)
Operations with flexibility but driving to Service Level Agreements and Performance Targets

Business Outcomes

- City Systems
  - Systems
  - Sensors & Devices
  - Data Sources
- Platform Services
  - Common integration middleware for systems, sensors, devices & networks
- Information Services
  - Common Information integration, management & abstraction services
- Interface Services
  - Interface abstraction & orchestration for applications, developers & technical users

Urban Systems Collaborative, April 2012
The key design decisions of the ISMP CORE Ontology for Operations

1. Represents messages, events and workitems as they flow through the system
2. Influenced by standards whenever possible (NIEM, CAP, NOAA, etc.)

3. Represents types of city services (not the city organization itself) so the administrative structure of a city can be assembled from ISMP building blocks
The key design decisions of the ISMP CORE Ontology for Operations

1. Represents messages, events and workitems as they flow through the system
2. Influenced by standards whenever possible (NIEM, CAP, NOAA, etc.)

3. Represents types of city services (not the city organization itself) so the administrative structure of a city can be assembled from ISMP building blocks
ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols.

```
<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KSTO1055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventCode>
      <valueName>SAME</valueName><value>SVR</value>
    </eventCode>
    <senderName>NATIONAL WEATHER SERVICE SACRAMENTO CA</senderName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
    </area>
  </info>
</alert>
```
ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols.

<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KSTO1055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventIdCode>
      <valueName>SAME</valueName>
      <value>SVR</value>
    </eventIdCode>
    <senderName>NATIONAL WEATHER SERVICE SACRAMENTO CA</senderName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
      <geocode><valueName>SAME</valueName><value>006109</value></geocode>
    </area>
  </info>
</alert>
ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols.

<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KSTO1055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventId>
      <valueName>SAME</valueName><value>SVR</value>
    </eventId>
    <senderName>NATIONAL WEATHER SERVICE SACRAMENTO CA</senderName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
      <geocode><valueName>SAME</valueName><value>006109</value></geocode>
    </area>
  </info>
</alert>
ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols

<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KSTO1055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventCode>
      <valueName>SAME</valueName><value>SVR</value>
    </eventCode>
    <sendName>NATIONAL WEATHER SERVICE SACRAMENTO CA</sendName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
      <geocode><valueName>SAME</valueName><value>006109</value></geocode>
    </area>
  </info>
</alert>
Urban Systems Collaborative, April 2012

ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols.

<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KS01055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventCode>
      <valueName>SAME</valueName>
      <value>SVR</value>
    </eventCode>
    <senderName>NATIONAL WEATHER SERVICE SACRAMENTO CA</senderName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
      <geocode><valueName>SAME</valueName><value>006109</value></geocode>
    </area>
  </info>
</alert>
ISMP represents what is known by one piece of software, which can be transmitted to a second piece of software, using a number of notification protocols.

<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>KSTO1055887203</identifier>
  <sender>KSTO@NWS.NOAA.GOV</sender>
  <sent>2003-06-17T14:57:00-07:00</sent>
  <status>Actual</status>
  <msgType>Alert</msgType>
  <scope>Public</scope>
  <info>
    <category>Met</category>
    <event>SEVERE THUNDERSTORM</event>
    <responseType>Shelter</responseType>
    <urgency>Immediate</urgency>
    <severity>Severe</severity>
    <certainty>Observed</certainty>
    <eventCode>
      <valueName>SAME</valueName><value>SVR</value>
    </eventCode>
    <senderName>NATIONAL WEATHER SERVICE SACRAMENTO CA</senderName>
    <headline>SEVERE THUNDERSTORM WARNING</headline>
    <area>
      <areaDesc>EXTREME NORTH CENTRAL TUOLUMNE COUNTY IN CALIFORNIA</areaDesc>
      <polygon>38.47,-120.14 38.34,-119.95 38.52,-119.74 38.62,-119.89 38.47,-120.14</polygon>
      <geocode><valueName>SAME</valueName><value>006109</value></geocode>
    </area>
  </info>
</alert>
<urgency>Immediate</urgency>
A Smarter City Semantic Model

A *semantic* data model of a city, if it is complete and authoritative, (1) simplifies the development of applications that require integrated access to city data sources and (2) enables solution reuse as we move from one city to the next.

Independently of using ETL for data consolidation, a *semantic* data model (3) can extend the metadata with new categories (SanitationServices, CrimesAgainstProperty) *without modifying the application or the data sources.*

A semantic data model for the city should assume that
1. Cities have their own data sources, not necessarily connected, and may not want to consolidate them.
2. Cities have non-standard organizations, departments and competencies.
Intelligent Semantic Model Palette - data model

ISMP is a non-normative, authoritative, modular, extensible semantic model for Smarter Cities, with a focus on What is happening in the city, and what parties are doing what, using what resources and assets and services.

It consists of a Core model that includes common classes (events and messages, stakeholders, departments, services, city landmarks and resources, KPIs, etc.), extensions by domain and customizations by city.

Simple language
• Classes + Inheritance + Relations + Inferencing
• Based on standards (OWL-QL, SPARQL)
• Mappable to UML
• Metadata annotations and Tagging

Authoritative
• Aligned with standards (CAP, NIEM, MISA/MRM, UCore)
• Validate with customer scenarios
• Validated with open city data

Intelligent Semantic Model Palette - data model

• Classes + Inheritance + Relations + Inferencing
• Based on standards (OWL-QL, SPARQL)
• Mappable to UML
• Metadata annotations and Tagging

Common building blocks

Organization/Operation profile

Features

AssetManagement
BuildingAndParcel
Transportation
Water
Weather

Additional features built by U.S.C. members?
Intelligent Semantic Model Palette tooling
ISMP is a non-normative, authoritative, modular, extensible semantic data model for Smarter Cities. It’s written using standard RDF/OWL editors (e.g. Protégé) and software (Jena)

The Scribe research effort is also
a. A modeling process (D-SCRIBE)
b. Tools to make the model usable. The first tool we’ve worked on, MIDO Data Access Service, allows the mapping of existing data to the ISMP model and is part of the process of customizing ISMP to a new city. Model update by Message Arrival is coming next (stay tuned).
Customizing Intelligent Semantic Data Palette in different cities

ISMP is NOT closed. We know that cities have different organizations, different service levels and different KPIs. The ISMP model is designed to provide the building blocks (service types, city departments, KPI taxonomies, CAP messages) that can be customized to define the overall operations of a city.
Research Philosophy
Principles

- **Make I.S.M.P. usable**
  - Assistance to real solution providers through flexible vocabulary (aliases, tagging) and context-awareness
  - Simplify understanding through model factoring, partitioning and perspectives (500−1000 concepts per perspective max)
  - Refer to open (defacto) standards if more context is needed, provide Business Glossary

- **Make I.S.M.P. Modular**
  - Weave together what is needed for a particular software package or customer
  - Have clear dimensions in our packaging, and patterns of assembly

- **Recognize the context of government data, “Policy-based Integration”, Resilience**
  - Many values may be returned as “unknown” or “unconfirmed” or “conflicting values stored”
  - Support passing a justification from the requesting program to the supplying program for auditing and for privacy support (tentative)
    - E.g. number of cars going into a military location may not be retrieved by all traffic optimization programs

- **Allow evolution of I.S.M.P. as understanding increases**
  - Not a static, rigid and fragile one-time deliverable
  - Show how model has evolved from previous version (2012 to 2013)

- **Make process of creation of I.S.M.P. participatory**
  - Core concepts are stable. Leverage open stds, and change when enough experience is gained (highest level of control)
  - Technology trends and best practices are modifiable
  - Project-specific additions are possible; learnings can feed more established concepts (lowest level of control)
  - Plan for this from day 1, invite additional people (IBMers & licensees) in, starting 3Q2012
SCENARIO
Scenario. Events in Washington D.C.

Suppose a Smarter City application that manages city operations wants to display citizen complaints (311 calls) on a map, filtered by a few user-defined constraints (times, locations, type of call, etc.)

A fraction of the 311 incident table (from DC Open Data) is below. Among the data is:

- Identifier
- Type of service (code + description)
- Time (ServiceOrderDate, ServiceResolution date, etc.)
- Place (Lon/Lat, Ward, PSA, District, etc.)
- The agency that should handle the request
- Various qualifiers (enum types): priority, resolution, etc.

Data Catalog

For years, the District of Columbia has provided public access to city operational data through the internet. Now the District provides citizens with the access to 483 datasets from multiple agencies, a catalyst ensuring agencies operate as more responsive, better performing organizations. Use the data catalog below to subscribe to a live data feed in Atom format and access data in XML, Text/CSV, KML or ESRI Shapefile formats. Please note that by accessing the data catalog and feeds, you agree to our Terms of Use. Please read before accessing the data. All data visualizations on maps should be considered approximate. The visualizations provided by this application include only records that can be mapped. If you have any questions or comments about the DC Data Catalog, please contact us at citydw@dc.gov. Visit DC FOIA page for information on FOIA requests.

<table>
<thead>
<tr>
<th>SERVICEREQUESTID</th>
<th>SERVICEPrio</th>
<th>SERVICETYPECODE</th>
<th>SERVICETYPESCODEDESCRIPTION</th>
<th>SERVICEORDERD</th>
<th>SERVICEORDAGENCY</th>
<th>AGENCYABBREVIATION</th>
<th>RESOLUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-000000078</td>
<td>UNKNOWN</td>
<td>DEPAHEAL</td>
<td>DOH</td>
<td>1/2/2010 7:51</td>
<td>OVERDUE OF DOH</td>
<td>DPW</td>
<td>Baited - fu</td>
</tr>
<tr>
<td>10-00000019</td>
<td>UNKNOWN</td>
<td>PARKENFO</td>
<td>Parking Enforcement</td>
<td>1/1/2010 11:47</td>
<td>CLOSED</td>
<td>DPW</td>
<td>Not Enforc</td>
</tr>
<tr>
<td>10-00000067</td>
<td>UNKNOWN</td>
<td>PARKENFO</td>
<td>Parking Enforcement</td>
<td>1/1/2010 21:01</td>
<td>CLOSED</td>
<td>DPW</td>
<td>Repeat Of</td>
</tr>
<tr>
<td>10-00000073</td>
<td>UNKNOWN</td>
<td>PRSVAVOP</td>
<td>Abandoned Vehicle Operations</td>
<td>1/1/2010 21:53</td>
<td>OVERDUE OF DPW</td>
<td>UNKNOW</td>
<td></td>
</tr>
<tr>
<td>10-00000049</td>
<td>UNKNOWN</td>
<td>SIGNS</td>
<td>Signs</td>
<td>1/1/2010 14:54</td>
<td>CLOSED</td>
<td>DDO</td>
<td></td>
</tr>
<tr>
<td>10-00000039</td>
<td>UNKNOWN</td>
<td>SISYINOD</td>
<td>SIOD</td>
<td>1/1/2010 13:41</td>
<td>OVERDUE OF DDOT</td>
<td>UNKNOW</td>
<td></td>
</tr>
<tr>
<td>10-00000098</td>
<td>UNKNOWN</td>
<td>SISYINOD</td>
<td>SIOD</td>
<td>1/2/2010 10:19</td>
<td>OVERDUE OF DDOT</td>
<td>UNKNOW</td>
<td></td>
</tr>
<tr>
<td>10-00000082</td>
<td>UNKNOWN</td>
<td>SNOW</td>
<td>Snow</td>
<td>1/2/2010 8:37</td>
<td>OVERDUE OF DPW</td>
<td>UNKNOW</td>
<td></td>
</tr>
</tbody>
</table>
Scenario. Events in Washington D.C.

The application may access directly this table by querying incidents according to given criteria:

```
A

SELECT SERVICEREQUESTID SERVICETYPECODE LATITUDE LONGITUDE WARD DISTRICT PSA DATEREPORTED FROM DC911 WHERE SomeConstraintHere
```

OR

The application may define an intermediate (data model) layer that:

- **B** Defines a `ServiceRequest` object that knows how to retrieve all the data from one or more tables.
- **C** Defines two objects, `ServiceRequest`, where all the common data to all service requests is, and `DC311SvceReq`, which captures the info specific to DC.

Notice that in (C), inheritance can be applied to locations (wards, districts, addresses, Lon/Lat points are ways to describe a location). Also, we could push further and have all kinds of abstractions, say, an event class that captures ID, Time, Location and Type.
Scenario. Events in Washington D.C.

Now suppose that the application wants to add crime incidents. The open data table is below. Notice that it looks similar to DC311, but not quite:

- ID’s have different format
- Time is ReportDateTime, and has a time
- Offenses do not have codes
- There’s no referring agency

So, from the point of view of the application:

**A** We can create another query for the DC911 table and consolidate the information at the application level (requires recompiling)

**B** We can add types and data to the object model, but this bloats the objects.

**C** We can use the inheritance hierarchy to refactor the information in the model. IF the model is well thought out, the changes are minimal... But we’ll need inferencing, infrastructure to keep the graphs...

We’ll be replicating RDF/OWL

<table>
<thead>
<tr>
<th>CCN</th>
<th>REPORTDATE</th>
<th>OFFENSE</th>
<th>BLOCKS</th>
<th>LATITUDE</th>
<th>LONGITUDE</th>
<th>WARD</th>
<th>DISTRICT</th>
<th>PSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>9185124</td>
<td>1/1/2010 0:00</td>
<td>THEFT F/AUTO</td>
<td>1900 B/O FENWICK ST N</td>
<td>38.91425947</td>
<td>-76.98444392</td>
<td>5</td>
<td>FIFTH</td>
<td>50</td>
</tr>
<tr>
<td>9185134</td>
<td>1/1/2010 0:00</td>
<td>ADW</td>
<td>800 B/O 21ST ST NE</td>
<td>38.90149128</td>
<td>-76.97417631</td>
<td>5</td>
<td>FIFTH</td>
<td>50</td>
</tr>
<tr>
<td>9185104</td>
<td>1/1/2010 0:00</td>
<td>ROBBERY</td>
<td>700 B/O MORTON ST NW</td>
<td>38.93190228</td>
<td>-77.02516974</td>
<td>1</td>
<td>THIRD</td>
<td>30</td>
</tr>
<tr>
<td>10000041</td>
<td>1/1/2010 0:00</td>
<td>ADW</td>
<td>2000 B/O GEORGIA AVE</td>
<td>38.91701078</td>
<td>-77.02189474</td>
<td>1</td>
<td>THIRD</td>
<td>30</td>
</tr>
<tr>
<td>10000086</td>
<td>1/1/2010 0:00</td>
<td>ROBBERY</td>
<td>600 B/O IRVING ST NW</td>
<td>38.92903929</td>
<td>-77.02213507</td>
<td>1</td>
<td>THIRD</td>
<td>30</td>
</tr>
<tr>
<td>10000115</td>
<td>1/1/2010 0:00</td>
<td>BURGLARY</td>
<td>3500 B/O STANTON RD</td>
<td>38.84368456</td>
<td>-76.97891927</td>
<td>8</td>
<td>SEVENTH</td>
<td>70</td>
</tr>
<tr>
<td>10000250</td>
<td>1/1/2010 0:00</td>
<td>ADW</td>
<td>1300 B/O OKIE ST NE</td>
<td>38.91421407</td>
<td>-76.98602973</td>
<td>5</td>
<td>FIFTH</td>
<td>50</td>
</tr>
<tr>
<td>10000366</td>
<td>1/1/2010 0:00</td>
<td>ADW</td>
<td>700 B/O 7TH ST NW</td>
<td>38.89913006</td>
<td>-77.02191663</td>
<td>2</td>
<td>FIRST</td>
<td>10</td>
</tr>
<tr>
<td>10000147</td>
<td>1/1/2010 0:00</td>
<td>ADW</td>
<td>1500 B/O MONROE ST N</td>
<td>38.93226905</td>
<td>-77.03560479</td>
<td>1</td>
<td>THIRD</td>
<td>30</td>
</tr>
<tr>
<td>10000141</td>
<td>1/1/2010 0:00</td>
<td>THEFT</td>
<td>2700 B/O O ST SE</td>
<td>38.87074185</td>
<td>-76.96837205</td>
<td>7</td>
<td>SIXTH</td>
<td>60</td>
</tr>
</tbody>
</table>
Scenario. Events in Washington D.C.

… And there are net benefits to a model-driven, semantic approach:

1. Applications can be coded ‘in the abstract’. E.g., Display all current events independently of whether they are 311 or 911.
2. Applications can refine the metadata without having to touch the code or the underlying data. E.g. Display all sanitation requests
3. Applications can be shielded from the details of the databases, like in the case of implicit joins. E.g. Display the names of the dispatchers associated with active requests.

In what follows, we’ll discuss the ISMP approach to semantic modeling to Smarter Cities data models through the Washington D.C. example.
Step 1. Customizing ISMP for Washington D.C.

Given that ISMP captures events, service types, date, etc. we had most of the model we needed, but not quite. For example, we hadn’t modeled all the services that the 311 table had.

ISMP is meant to provide the building blocks to define the elements of a Smarter City, and so it’s designed to be customized and extended. So we created a new file for DC (with namespace ‘WDC’ that imports the ISMP Core Palette.

We may want to customize ISMP for a variety of reasons:

- **SuperCans** is a DC-specific program and it will likely remain in the DC specific classes.
- **Collecting Illegal Dumping** or Seasonal Collection were not contemplated in the core, and they may be marked for promotion at a later date (using the modelPromotion annotation)
- **Criminal Incident911** and ServiceRequest311, are made specific for Washington DC. This is a design decision, in case we need to show/coordinate with other cities.
Step 2. Mapping instance data to the model

Next, we map the data in the columns to either a data property (transferring the data into that data property, like in the case of SIMPLEREQUESTID) OR a class (to match an enumerated type, which in the case of ISMP is represented as a taxonomy of classes.)

This mapping is done through a model and tool called MIDO.

The (MIDO) mapping is done once at design time and it’s done manually, but it can also be done semi-automatically (currently not implemented).

We’ll cover the details of the mapping later. For now, let’s assume that the columns in the two tables have already been mapped.
MIDO Mapping
MIDO imports instance data into the ISMP using the mapping. The data can be queried now as if it was native to the ontology... This means that inferencing is automatically used.

<table>
<thead>
<tr>
<th>ServiceRequestID</th>
<th>ServiceType</th>
<th>dateOrdered</th>
<th>Status1</th>
<th>Status2</th>
<th>District</th>
<th>Ward</th>
<th>latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-000000099</td>
<td>DCBulkCollection</td>
<td>Date1/2/201010:19</td>
<td>CLOSED</td>
<td>-</td>
<td>SEVENTH</td>
<td>8</td>
<td>38.83144</td>
</tr>
<tr>
<td>10-00000004</td>
<td>DCRcycleContainerCollection</td>
<td>Date1/2/201010:13</td>
<td>CLOSED</td>
<td>OVERDUE</td>
<td>SECOND</td>
<td>2</td>
<td>38.90752</td>
</tr>
<tr>
<td>10-000000096</td>
<td>DCBulkCollection</td>
<td>Date1/2/201010:16</td>
<td>CLOSED</td>
<td>-</td>
<td>THIRD</td>
<td>1</td>
<td>38.92552</td>
</tr>
<tr>
<td>10-000000084</td>
<td>DCTrashCollection</td>
<td>Date1/2/20108:50</td>
<td>CLOSED</td>
<td>OVERDUE</td>
<td>SIXTH</td>
<td>7</td>
<td>38.86723</td>
</tr>
<tr>
<td>10-000000088</td>
<td>DCSanitationEnforcement</td>
<td>Date1/2/20109:46</td>
<td>CLOSED</td>
<td>OVERDUE</td>
<td>-</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>10-000000085</td>
<td>DCRcycling</td>
<td>Date1/2/20108:51</td>
<td>CLOSED</td>
<td>OVERDUE</td>
<td>SIXTH</td>
<td>7</td>
<td>38.86723</td>
</tr>
<tr>
<td>10-000000079</td>
<td>DCSuperCans</td>
<td>Date1/2/20108:11</td>
<td>CLOSED</td>
<td>-</td>
<td>FIFTH</td>
<td>5</td>
<td>38.91151</td>
</tr>
<tr>
<td>10-000000082</td>
<td>DCSnowRemoval</td>
<td>Date1/2/20108:37</td>
<td>OPEN</td>
<td>OVERDUE</td>
<td>SIXTH</td>
<td>7</td>
<td>38.90212</td>
</tr>
</tbody>
</table>

(DC311 Mapping)

The ISMP model annotates some services as Sanitation Service.
The current prototype implementation
Step 3. Query through the model. Query abstract classes

The data from DC Service Requests and Crime incidents can now be queried together as events, not just as service requests or criminal incidents.

<table>
<thead>
<tr>
<th>Row</th>
<th>Event</th>
<th>descriptorLabel</th>
<th>District</th>
<th>Ward</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10000716</td>
<td>STOLEN AUTO</td>
<td>SEVENTH</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>10000328</td>
<td>THEFT F/AUTO</td>
<td>FIFTH</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>10000315</td>
<td>THEFT F/AUTO</td>
<td>THIRD</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>10000672</td>
<td>THEFT</td>
<td>THIRD</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>10000713</td>
<td>STOLEN AUTO</td>
<td>THIRD</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>10000665</td>
<td>ROBBERY</td>
<td>SIXTH</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>10000200</td>
<td>STOLEN AUTO</td>
<td>SIXTH</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>10000237</td>
<td>THEFT</td>
<td>SECOND</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>10000353</td>
<td>THEFT</td>
<td>SECOND</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>10000147</td>
<td>ADW</td>
<td>THIRD</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>10000580</td>
<td>THEFT F/AUTO</td>
<td>THIRD</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>10000055</td>
<td>THEFT</td>
<td>SECOND</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>10000250</td>
<td>BURGLARY</td>
<td>SECOND</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>10000058</td>
<td>ROBBERY</td>
<td>SEVENTH</td>
<td>8</td>
</tr>
</tbody>
</table>

Query: All Events in DC, with type, District and Ward

```
WHERE {
  ?Event Scribe:hasEventDescriptor ?x .
  OPTIONAL {
    ?Event Scribe:hasEventDescriptor ?descriptor.
    ?descriptor Scribe:IncidentType .
  }.
  OPTIONAL {
    ?district a Scribe:PoliceDistrict .
  }.
  OPTIONAL {
    ?Event Scribe:eventLocatedIn ?ward .
  }.
}
```

Notice that some of the data is missing in the original table… That’s still ok
Step 3. Query through the model. Annotation Metadata

As shown previously. The inferencing in the ontology can be leveraged in a query.

**Query: Public Sanitation Service Requests**

```
SELECT * WHERE {
    ?subject Scribe:requestAssociatedToServiceType ?serviceType .
    ?serviceType a Scribe:SanitationService .
}
```

Requests:

<table>
<thead>
<tr>
<th>ServiceRequestID</th>
<th>ServiceType</th>
<th>dateOrdered</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-00000099</td>
<td>DCBulkCollection</td>
<td>12/20 10:19:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000094</td>
<td>DCRcycleContainerCollection</td>
<td>12/20 10:13:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000096</td>
<td>DCBulkCollection</td>
<td>12/20 10:16:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000084</td>
<td>DCTrashCollection</td>
<td>12/20 08:50:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000088</td>
<td>DCSanitationEnforcement</td>
<td>12/20 10:46:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000085</td>
<td>DCRcycling</td>
<td>12/20 08:51:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000079</td>
<td>DCSuperCans</td>
<td>12/20 08:11:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000082</td>
<td>DCSnowRemoval</td>
<td>12/20 08:37:00</td>
<td>OPEN</td>
</tr>
<tr>
<td>10-00000090</td>
<td>DCBulkCollection</td>
<td>12/20 09:59:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000087</td>
<td>DCBulkCollection</td>
<td>12/20 09:37:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000081</td>
<td>DCTrashCollection</td>
<td>12/20 08:36:00</td>
<td>CLOS</td>
</tr>
<tr>
<td>10-00000089</td>
<td>DCBulkCollection</td>
<td>12/20 09:50:00</td>
<td>CLOS</td>
</tr>
</tbody>
</table>
Step 3. Query through the model. Implicit join

Everything in a semantic model is connected. The service request can be linked to the name of the dispatcher of the department.

**Query:** Select events associated to dept of Public Works and his dispatcher

```
SELECT DISTINCT ?event ?department ?dispatcherName
WHERE {
  ?event a Scribe:ServiceRequest .
  ?event Scribe:requestHandledBy WDC:DepartmentOfPublicWorks .
  ?event Scribe:requestHandledBy ?department .
  ?department Scribe:associatedTo ?dispatcher .
  ?dispatcher a Scribe:Dispatcher .
  ?dispatcher Scribe:roleOfPerson ?dispatcherName .
}
```
Engagement Customization and Asset Maturation

I.S.M.P. at the time work began on engagement with City X

Additions from City X which were added to I.S.M.P.

Metadata Updates (X)
Engagement Customization and Asset Maturation

<table>
<thead>
<tr>
<th>New I.S.M.P. after City Y</th>
<th>I.S.M.P. at the time work began on engagement with City X</th>
<th>Additions from City X which were added to I.S.M.P.</th>
<th>Additions from City Y which were added to I.S.M.P.</th>
</tr>
</thead>
</table>

| Metadata Updates (X,Y)   |                                                        |                                                  |                                                  |
Engagement Customization and Asset Maturation

<table>
<thead>
<tr>
<th>New I.S.M.P. after City Z</th>
<th>Additions from City X which were added to I.S.M.P.</th>
<th>Additions from City Y which were added to I.S.M.P.</th>
<th>Additions from City Z which were added to I.S.M.P.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.S.M.P. at the time work began on engagement with City X</td>
<td>Metadata Updates (X, Y, Z)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Engagement Customization and Asset Maturation

I.S.M.P. at the time work begins on engagement with City X

Not Needed for City X But already in I.S.M.P.

Additions for City X which will be added to I.S.M.P. for Reuse

Additions for City X which will Not be added to I.S.M.P.

Additions from City Y which were added to I.S.M.P.

Additions from City Z which were added to I.S.M.P.

New I.S.M.P. after City Z

Metadata Updates (X, Y, Z)

Additions for City X not added

Additions for City Y not added

Additions for City Z not added

Delivery To City X
CityForward.org – IBM’s experimentation with data sharing with the numerate, literative “person in the city”

CPS: almost as many students drop out as graduate each year

Chicago, IL (MSA), USA - CITY OF CHICAGO SCHOOL DISTRICT 299 (City)

<table>
<thead>
<tr>
<th>Year</th>
<th>Diploma Count</th>
<th>Dropout Count, Grades 9-12</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>18,000</td>
<td>10,000</td>
</tr>
<tr>
<td>2007</td>
<td>22,000</td>
<td>12,000</td>
</tr>
<tr>
<td>2008</td>
<td>20,000</td>
<td>14,000</td>
</tr>
</tbody>
</table>
SCRIBE enables consistent cross-domain solutions for Smarter Cities

- Predictive Modeling
  - Orchestration
  - High
  - Packaged
  - Low
  - None

- Flexible, semantics-driven
  - Multi-Domain

- Cross-domain Data/Semantics
  - ISMP (2013)
  - Traffic into Hub
  - Intelligent Operations Ctr (leverages ISMP)

- Single-Domain
  - Traffic
  - Water Infrastructure
  - Public Safety
  - Video Surveillance

- Integrated
  - Traffic
  - Water Infrastructure
  - Public Safety
  - Video Surveillance

- Unified
  - Traffic
  - Water Infrastructure
  - Public Safety
  - Video Surveillance
What is ISMP?

- a data model
- a semantic data model
- a modular semantic data model
  - by domain
  - by different cities
  - by dimensions

```
CityFinanceAndBudgetService
CityOperationsService
  - ElectricityManagementService
  - FireService
  - ParksAndRecreationService
  - PoliceService
  - SanitationService
  - TelephoneAndTelecommunicationsService
TransportationAndTrafficService
  - PublicTransportationService
    - CharteredPublicTransportationService
    - PublicBusService
    - RailService
  - TrafficService
    - RoadMaintenanceAndRepair
    - TrafficManagementService
CityPlanningAndDevelopmentService
```
Business Value: Decision-making optimized!
### Example scenarios for cross-domain decision support
(Intelligent Operations Center offering can enable)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Domains / Factors</th>
<th>Target Outcome</th>
<th>Additional KPIs</th>
<th>Data Sources / Predictive Models</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emergency Response</td>
<td>Traffic, people, public transportation</td>
<td>Meet emergency response time requirements</td>
<td>Safety, cascading effects such as impact on travel</td>
<td>Vehicle, subway, bus, pedestrian traffic data; traffic simulation, GIS</td>
</tr>
<tr>
<td>Flooding, Overflow</td>
<td>Weather, sewage, traffic</td>
<td>Maximize flood clearing (free blockages, increase water carriage), minimize damage</td>
<td>Minimize impact to transportation, sewage</td>
<td>Weather, flood and damage prediction, traffic prediction, GIS</td>
</tr>
<tr>
<td>Power Restoration / Brownout</td>
<td>Weather (storm damage, heat wave), power, city services</td>
<td>Minimize restoration time/optimize restoration sequence</td>
<td>Ensure continuity of social services (school meals for kids)</td>
<td>City service requirements, weather, power grid</td>
</tr>
<tr>
<td>Infrastructure Repair</td>
<td>Transportation, multiple infrastructure</td>
<td>Maximize repair/maintenance, minimize time/cost</td>
<td>Maximize co-located predictive maintenance, minimize impact to transportation</td>
<td>Water, electrical, gas, sewage, roads data/risks, traffic data/prediction</td>
</tr>
<tr>
<td>Police Dispatch</td>
<td>Public safety and security, traffic</td>
<td>Meet response time requirements</td>
<td>Ensure overall safety and security</td>
<td>Real-time location data, traffic, crime data, cameras</td>
</tr>
</tbody>
</table>
Technical Vision – Why Semantics?

An *architected* semantic model can make our Smarter City solutions reusable and profitable, simplifying the transfer of knowledge across projects

- Reuse of common model patterns reduces errors and increases quality
- Flexible searching and browsing makes it usable to developers and domain experts
- Semantic data mappings simplify/strengthen data integration

**Database Schema**
Brittle, limited language, Data organization is fixed and visible to user

**Intelligent Semantic Model Palette (RDF/XML)**
Architected approach to industrial-strength semantic models
- Componentization
- Patterns
- Community authoring/verification

**RDF/OWL**
Lacks structure hard to understand/verify scale for industrial applications
Reducing every agency’s software to understanding at most 2 representations, not \( n \) representations.
Information serializations logically will be “RDF” (Linked Data), Data Exchange logically will be “XML”

**Why this is a research problem**
- The state-of-the-art for composable models has each model only address one “level” – business strategy and objectives level, or business logical operating design level, or software solution macro component level . . . And even then, models with over 150 entities have proven “fragile” (resistant to organic growth) and “sluggish” (takes quite a bit of time for developers to learn)
- The state-of-the-art for model production depends upon largely complete requirements and scenarios at the beginning, but in Smarter Cities, we need to be able to accept new requirements at all stages of the model develop and validation process
## Technical Vision and Research Challenges – p3

### Methods

<table>
<thead>
<tr>
<th>Collaborative, Governed Authoring</th>
<th>Terms familiar to stakeholders</th>
<th>Modular, Well-defined Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context-Based Authoring (where to 'insert' Concept C?)</td>
<td>Help in Browsing/Searching (Are there other related concepts? E.g. Everything related To Transportation)</td>
<td>Flexible Data Mappings (data vocabulary and organization)</td>
</tr>
<tr>
<td>Tags to help in authoring</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Iterative Design

- Scribe Models
- Scribe Tools

### Smarter Cities Applications

- Urban Information Layer Work with Urban Planners
- Pre-IIC apps&data
- IIC apps Starting
- Deep Thunder (Lloyd Treinish)
- S3 (Chiao-fe Shu)
3 ways to architect a family of solutions

No uniform data model

- Every program inputs, outputs, and holds the optimal representation for it
- A great deal of mapping, transform, ETL, and duplication of data in databases is used

Development cost is: Back-end Heavy

Uniform data model (proven)

- Common representations
- Relationships between data entities have 1 of 3 costs to traverse:
  - Very low (exmpl: child node of this node in XML)
  - Medium (ex: join using foreign key)
  - High (search, views)
- Exmpl Notation: UML

Development cost is: Spread throughout

Semantic data model (emerging)

- Common representations
- Relationships between data entities have 1 of 2 costs to traverse:
  - Very low (example: direct relation)
  - Normal (example: reasoning)
- Example Notation: OWL

Development cost is: Front-end Loaded
City Customization Tool – a vision

Inputs

Representation of City Accessible Data Repositories & Feeds, & City Physical & Logical Organization

ISMP.domain

Outputs

ISMP.domain

ISMP.domain.city

Instance Access Metadata For IOC Model Server

How do we explain Whether there will be a cost to include something the city does not now use, or no cost?
City Customization Tool with multiple sessions

Inputs

Representaton of City Accessible Data Repositories & Feeds

ISMP.domain

Open Questions List
(will be empty at the beginning of the 1st session)

Outputs

ISMP.domain

Instance Access Metadata For IOC Model Server

ISMP.domain.city

Open Questions List
(will be empty at the end of the final session)
IOC City Customization Tool with multiple sessions

Inputs

- Agency-by-Agency Ability To Publish and Receive CAP Alerts, Workorders, etc.
- Representation of City Accessible Data Repositories & Feeds
- Scribe.domain2
- Scribe.domain1
- Open Questions List (will be empty at the beginning of the 1st session)

Outputs

- IOC Configuration
- Scribe.domain1.city
- Scribe.domain2.city
- Instance Access Metadata For IOC Model Server
- Open Questions List (will be empty at the end of the final session)
How. Our process – D-SCRIBE -- Part 1

1. Collect
   - City Data
     - Open Innovation Platform, Yellow Button...
   - EXDL
   - WS-Calendar
   - Open311
   - CAP
   - NIEM
   - OpenStreetMaps
   - IBM Intelligent Buildings
   - MISA
     - Municipal Ref Model
   - Water Management
     - Neighborhood Re-development
   - Business Impr’nt District

2. Describe Scenarios
   - Environment
   - Traffic
   - Weather
   - Water
   - Flooding

3. Select Terms
   - Add Definitions to InfoSphere Business Glossary
     - Judgment Expertise

4. Semantic Queries - Abstract
   - Example Queries
   - KPI Formulas

5. Model Semantic Relations, Review
   - What should be added to CORE?

6. Modularize, Metadata
   - SCRIBE.geo
   - SCRIBE.time
   - SCRIBE.event
   - SCRIBE.asset
   - SCRIBE.phys_city
   - SCRIBE.traffic
   - SCRIBE.environment

7. Example Instance Data
   - SCRIBE.phys_city.data.Dublin
   - SCRIBE.phys_city,MIDO.Dublin

8. Select Aspects, Generate (SOA Resources)
   - WSDL
   - ER
   - XSD

9. Write & test Example Queries
   - SCRIBE.TestQueries.Traffic.Dublin

1. Continue looking for Standards
How. Our process – D-SCRIBE -- Part 2

10. Connect to Sources “at Rest” (do – 1/3)
   - Dublin GIS City Data (via STAT)
     - Dublin Traffic Observation Sensors
     - Dublin Weather Observation Sensors
     - Dublin Public Safety Process Mgr
     - Dublin Intelligent Buildings Gateway
     - Dublin Water Delivery Gateway

11. Design Enriched CAP
   - Environment
   - Traffic
   - Weather
   - Water
   - Flooding

12. Coach Developers
   - Java Code
   - Data Model Extension

13. Coach Customizers
   - Policy-Governed Controlled Cross-Agency Visibility


15. Apply to Reports
   - Define Reports
   - Clarify Values in Reports (using BG)

16. Apply to Analytics
   - Predict/Optimize models

17. Scenarios Change as City Adapts to Culture of Data
   - Additional Sensors Installed
   - Decisions Informed by Predictive Analytics

Return to Step 2

SCRIBE.phys_city,MIDO.Dublin
How. Our process – D-SCRIBE -- Part 1.1

1. Collect

City Data (Open Innovation Platform, Open Data on the Web)
- EXDL
- WS-Calendar
- Open311
- CAP
- NIEM
- OpenStreetMaps

IBM Intelligent Buildings

Water Management

MISA Municipal Ref Model

2. Describe Scenarios

Environment

Traffic

Weather

Water

Flooding

3. Select Terms

4. Semantic Queries - Abstract

Neighborhood Re-development

Business Improvement District

5. Model Semantic Relations, Review

6. Modularize, Generate (SOA Resources)

7. Example Instance Data

8. Select Aspects, Generate (SOA Resources)

9. Write & test Example Queries
How. Our process – D-SCRIBE -- Part 1.2

1. Collect
2. Describe Scenarios
3. Select Terms

Add Definitions to InfoSphere Business Glossary

4. Semantic Queries - Abstract
   - Example Queries
   - KPI Formulas

1. Continue looking for Standards
How. Our process – D-SCRIBE -- Part 1.3

5. Model Semantic Relations, Review

6. Modularize, Metadata

8. Select Aspects, Generate (SOA Resources)

WSDL
ER
XSD

9. Write & test Example Queries

SCRIBE.TestQueries.Traffic.A-city

What should be added to CORE?

7. Example Instance Data

SCRIBE.phys_city,data.A-city
SCRIBE.phys_city,MIDO.A-city
How. Our process – D-SCRIBE -- Part 2.1

10. Connect to Sources “at Rest” (do ~ 1/3)
11. Design Enriched CAP
12. Coach Developers
13. Coach Customizers
15. Apply to Reports
16. Apply to Analytics
17. Scenarios Change as City Adapts to Culture of Data

- A-city GIS City Data (via STAT)
- Environment
- Traffic
- Weather
- Water
- Flooding
- A-city Traffic Observation Sensors
- A-city Weather Observation Sensors
- A-city Public Safety Process Mgr
- A-city Intelligent Buildings Gateway
- A-city Water Delivery Gateway
- SCRIBE.phys_city.MIDO.A-city

Return to Step 2

Urban Systems Collaborative, April 2012
10. Connect to Sources “at Rest” (do – 1/3)
11. Design Enriched CAP
12. Coach Developers
13. Coach Customizers
15. Apply to Reports
16. Apply to Analytics


Java Code
Data Model Extension
Policy-Governed Controlled Cross-Agency Visibility
Define Reports
Clarify Values in Reports (using BG)

Predict/Optimize models

17. Scenarios Change as City Adapts to Culture of Data

Additional Sensors Installed
New Citizen-Supplied Data
Decisions Informed by Predictive Analytics

Return to Step 2
Questions we want to Answer In Year 2011-12

- How to harmonize Overlapping Standards
- Plug-and-play Of different domain Ontologies (JEC – “just enough coupling”)
- Adaptation to one city is Straightforward and maintainable (use DC Public Events then Singapore(?) as testcase)
- Evidence Driven Prioritized Ontology Evolution
- Representation of Abstract Processes
- A Data Access Service (for IOC) that provides Ontology Views (MIDO), migrateable in 2012 to leverage DB2 10 NoSQL Graph Store
- Java Programmers Interface (JENA based)
Standardization: DC and San Francisco Open Data Sets
And Reuse: in Rio (Brazil) and HCMC (Vietnam) where 311 may not exist

SF 311 Calls
- Graffiti Complaint – Private Property
- Street and Sidewalk Cleaning
- Sewer Issues
- Street Defect
- Tree Maintenance or Damage
- Illegal Postings

DC 311 Calls
- Snow/Ice Removal
- Streetlight Repair
- Abandoned Vehicle
- Traffic Signal Maintenance
- Tree Inspection
- Tree Pruning
- Tree Removal

311 Event
- Sanitation Enforcement
- Municipal Code Violation
- Infrastructure Maintenance
- Urban Forestry

ETL
- Type checking
- Data Transformation
- Vocabulary integration

ISMP MODEL
- Semantic categorization & interpretation (scope)
- (Analytic) Services Mapping

Input
- Urban Forestry Call District

Correlate 311 call types

Probability of Sidewalk Maintenance

Database Schemas, Queries, Services (WSDLs)
### SUMMARY

<table>
<thead>
<tr>
<th>What</th>
<th>Who</th>
<th>Why</th>
<th>When</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tools</strong>&lt;br&gt;To Build Logl Models To Consume Models To Discuss Models To Compose Models To Search Models To Navigate Models</td>
<td></td>
<td>Attack limitations of current Tiny-team of architect approach (Web 2.0 means modeling leveraging aspect-oriented design)</td>
<td>Core + Weather – May 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Extensible Semantics Harmonization Refactoring Leverage Standards</td>
<td>Transport, Sensors, KPIs, - August</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tools for Consumers Nov 2012</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>And on-and-on</td>
</tr>
</tbody>
</table>
Describe how city coordination and optimization can be done with software and data using the composable ISMP logical models.

Support year-by-year flexible improvement in integration.
Research 2011-13 roadmap for O.I.P. exploiting ISMP

Collaborative data publication, access & sharing with manual anonymization
- Initial City Content, 90 data sets from Dublin, growing to 250 (plan to grow to 1200 by March 2012)

Continuous data collection from mobile & sensor
- Ingestion of volumes of continuous data (start with vehicle location)
- Prototype of spatio-temporal semantics across domains

Prototype tools for user mash-up exploiting semantic research results
- Capture semantics of external, noisy & social data sources
- Citizen engagement
- Continuous sensor & mobile data feeds
- External & Value-added Sources
- Factual & statistical city data

Research Challenges
- Simplify scalable privacy and security of resources
- Automate assimilation and sharing of resources
- Robust models to organize and represent resources and their context
  (SCRIBE - 2012-2H)
- Investigate intuitive ways to compose resources for development, mash-up & visualization
  (SCRIBE, Citizen Engagement 2013-1H)