Data: get big ... but get smart

Han Wammes makes the case that only Smart Data will help a Smart Society thrive on Big Data

It is interesting to see how Smart City initiatives, initially driven from a technology perspective and with their focus on the Internet of Everything and Big Data, are evolving into initiatives that are creating value to serve sustainable, economical, inclusive and environmental developments - the real purpose of Smart City initiatives. However, as the term Smart City suggests, Smart City projects stop at city boundaries. The term Smart Society was introduced to circumvent this issue.

Smart Data versus Big Data
Although Big Data technologies are primarily about how to deal with the volume and velocity of data coming from sensors, mobile devices and social networks, Big Data is still data like any other. Unless we add meaning (semantics) to data, it remains meaningless, so we need to be able to understand it. Likewise if we don't have a purpose for the data, why collect it in the first place? The same reasoning applies to Data Quality; quality of data defines the reliability of the decisions we make based on this data, so it needs to be fit for purpose.

These aspects I call Smart Data. And while some technologies can do a lot - Data Analytics, for example, as the next big thing - these ultimately also require Smart Data.

Big data as a concept should be defined around five aspects: data volume, data velocity, data variety, data veracity and data value. Of these characteristics, only the volume and velocity aspects refer to the data generation process and how to capture and store the data. The veracity and value aspects deal with the quality and the usefulness of the data, while the variety aspect deals with the diversity of data.

Smart Data for a Smart Society
All these aspects of Smart Data are not new; the lack of Semantics and Governance of Data Quality has caused many e-Government projects to fail, and despite unifying strategies such as Open Standards-based Cloud platforms from different technology vendors, a lot of data silos still exist. So we should be careful not to fall into the same trap as we did in the past.

Why is the current situation different then? There are several reasons for that:

- Boundary-less Information Flow™ (Open Group); there are now virtually no boundaries to stop making information freely available to anyone

Dutch Geospatial Information Models (Adapted from www.geonovum.nl)

Semantic Web Stack (Adapted from www.w3.org)

- Technology developments (Quantum Cryptography) can now provide the required security
- The call for data transparency; government policies require the opening up of data (both as is or anonymously)
- Open Data needs to be Smart, thus requires Semantics and Governance of Data Quality
- Linked Data (W3C) especially when it is Open (LOD), not only simplifies the way to access data in a unified way, but will likely lead to better enterprise interoperability
- There is a societal need to change the way we collaborate so everyone can participate in this digital transformation to a Smart Society
- A Smart Society can only thrive on Smart Data
- Location and time provide context, while logic provides knowledge

As mentioned before, data in itself is meaningless. Semantics are needed to be able to answer questions such as 'who' and 'what', while location and time can provide the context needed to answer the 'where' and 'when' questions. Thus semantics makes data meaningful especially when location and time are included as context. Eventually when adding logic, as we can find in ontologies (e.g. OWL) and rules engines, we will also be able to answer the 'why' questions. Thus, semantics and logic-based Smart Data form the basis of what we would call Data Analytics. They also provide the means to better govern the data management and data quality management process through automation.

Creating insight

It is argued that the discussion should be about Smart Data as an integral part of Big Data. As Smart Data is about creating insight, it is imperative that we understand data: that it is fit for purpose; that we reduce the noise in data, and that data is actionable. In particular, location data combined with temporal data, helps create the context needed to make better logic-based decisions that help answer the 'why' questions.

By combining semantics and rules and taking a holistic approach towards Smart Data in standards-based, platform agnostic, open technology components it is possible to create a fully automated governed data lifecycle management environment. Such an environment allows for continuous data quality improvement through data cleaning, data integration, data inference (data mining) and data harmonisation[1].

By including location and time, facts are given the appropriate context to answer all questions, from who to what and from where to when. Adding logic gives us the ability to intelligently answer why, thus providing the perfect basis for Data Analytics.

Semantics is about the meaning of data. So it is necessary that
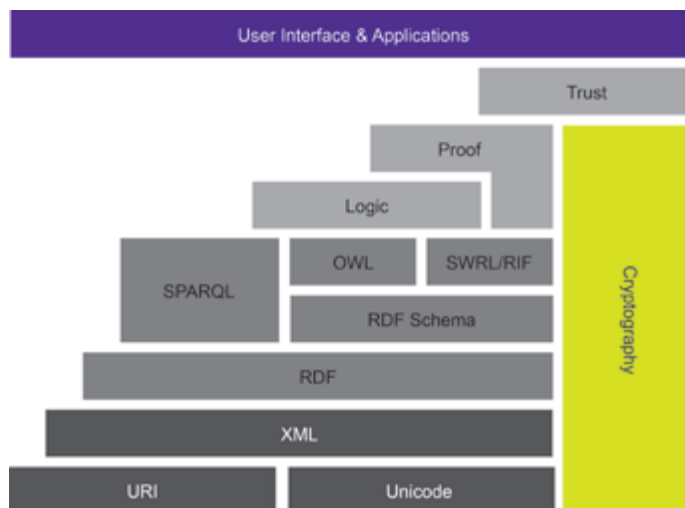
people, organisations and legislators agree on that meaning. Fortunately, many standardisation efforts are on the way, not only from well-known globally operating standardisation organisations such as OASIS, ISO and W3C to name a few, but also from IT domains (e.g. OGC in the geospatial domain) as well as from local initiatives such as that initiated in the Netherlands around Information Modelling and the Smart City Conceptual Model. In the UK, BSI has evolved its PAS182 data concept model[2] to provide a similar high-level information model of a Smart City.

Dutch treat

Since I live and work in the Netherlands, let's look at how semantics and base data, including geospatial data, go hand-in-hand in the effort to create a sufficient knowledge base for semantic interoperability in a collaborative way by the public, private and academic sectors. This triple helix collaboration should ensure that the Netherlands takes a unified approach towards data - from an enterprise perspective to data-sharing to anyone in a Smart Society context. As the geospatial domain is currently the most advanced, I will focus on this domain.

A couple of years ago it was decided that all geospatial data relevant to the natural and built-up environment should be made generally available as information models with their focus on data exchange and compliance with International, European, National, Sector and Local standards (e.g. XML Schema and GML for the geospatial part, modelled in UML).

In practice, it means that interoperability is required for at least the most common denominator between all consumers of this data. The Dutch Cadastre and Geonovum[3] (the Dutch geospatial standardisation institute) play key roles here, as do other relevant players ranging from public safety organisations to utility companies. The public sector organisations that maintain authoritative data have been defining key registers to establish which information models (based on XML Schema (XSD) but modelled in UML) partly overlap with the earlier-mentioned information models.

Although not yet finished, the implementation and harmonisation of these information models should be complete by 2017.

In the meantime, Geonovum is heavily involved in a national Linked Open Data Initiative[4] (Platform Linked Open Data Nederland). This is looking into how the existing information models can be published as Linked Open Data involving Semantic Web standards such as OWL, RDF, SPARQL, SKOS and JSON-LD.

The impact on Smart Society

I recently sat as a jury member for a contest organised by Smart-DataCity that was charged with selecting the most Smart City ("De Slimste Binnenstad"), and rating several Smart City initiatives in the
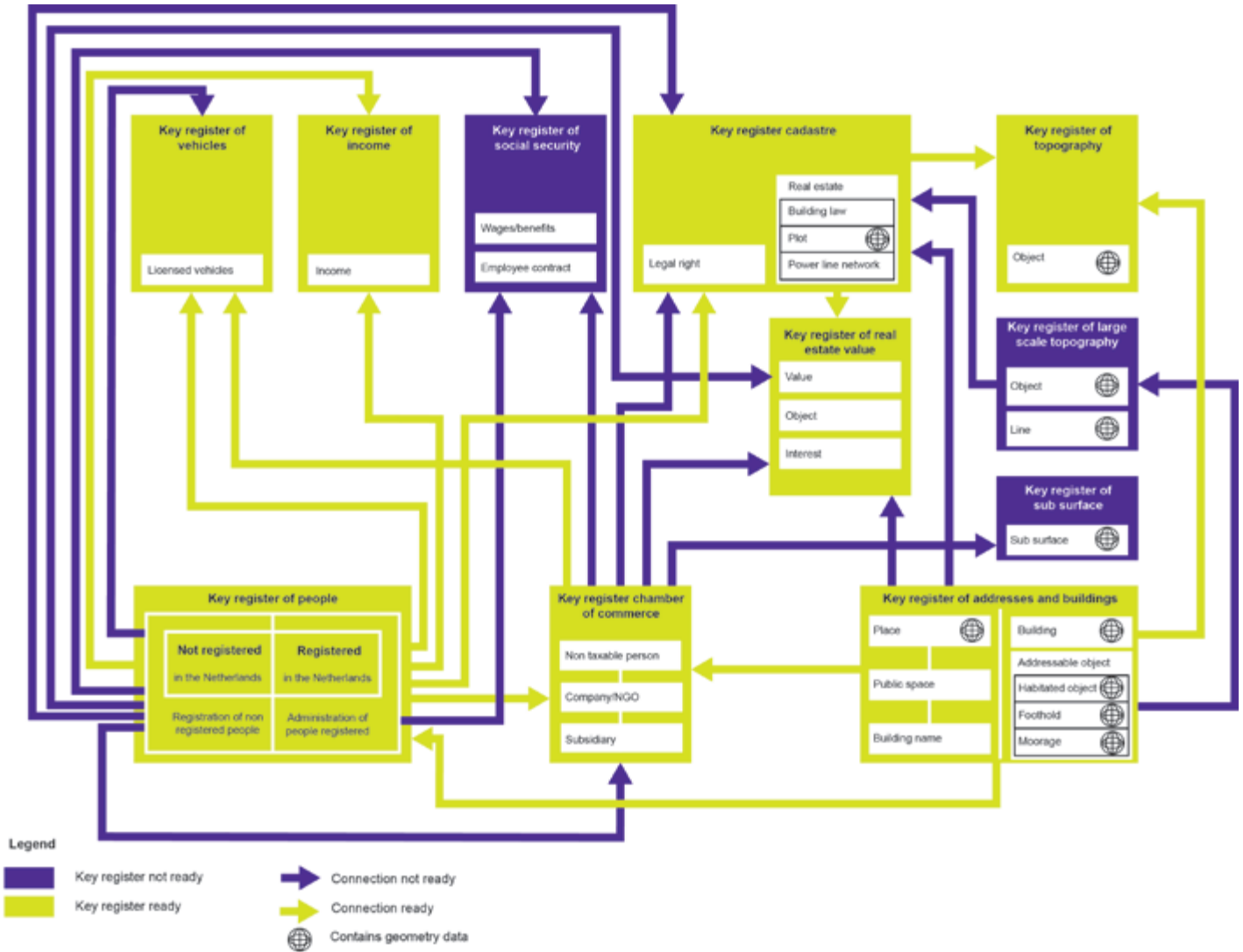
Dutch Key Registers (Adapted from www.digitaleoverheid.nl)

Netherlands. The criteria for the contest was based on the Smart Cities Maturity Model and Self-Assessment Tool of the Scottish Cities Alliance5.
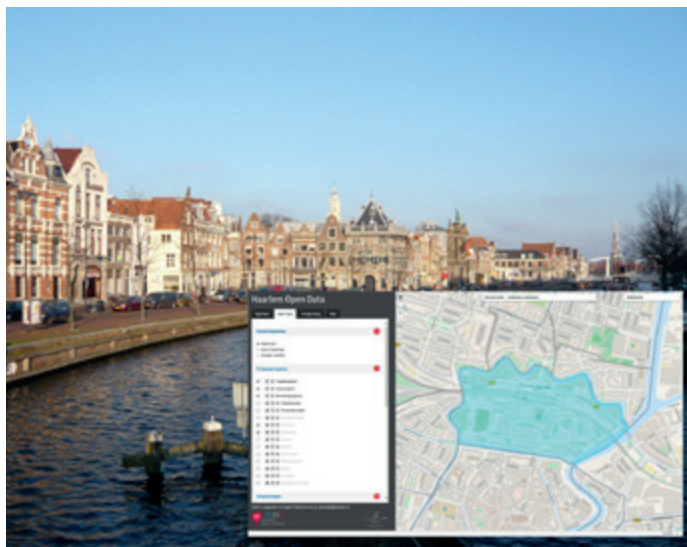
Although not a winner in a nation-wide smart city contest, only Haarlem in The Netherlands was able to produce a real Open Data platform that allowed for the active participation of citizens. *Photo:* Wikipedia. *Screen shot:* Vicrea Smart Solutions

What I found was that, although most cities are pretty aware of what needs to be done in terms of data strategy, its practical implementation is a quite different matter. Although some good examples exist, ranging from water management to developing a 100% sustainable micro grid, only one project was able to showcase a truly Smart City initiative. Here, the City of Haarlem was able to produce a real open data platform allowing for the active participation of citizens. It didn't win on this occasion, but showed us that the biggest challenge for any organisation is to create a genuinely open data platform.

References
1. David Loshin (2001) *Enterprise Knowledge Management: The Data Quality Approach.* 0-12-455840-2
2. http://www.bsigroup.com/en-GB/smart-cities/Smart-Cities-Standards-and-Publication/PAS-182-smart-cities-data-concept-model/
3. www.geonovum.nl
4. www.pilod.nl/wiki/ELD_Project_Summary
5. www.smartdatacity.org

Hans Wammes is as Business Development Manager with 1Spatial (www.1spatial.com) and is responsible for setting up a Smart Society value proposition. His experience is in both GIS and IT, specifically dealing with Information Management.