Ensuring Quality in an "Internet of Things"

Messages between devices, or to/from humans benefits from structure.

Version 4, February 5, 2018
Prepared by Michael Scofield, M.B.A.

Synopsis

The "internet of things" (which means connectivity between devices) requires that messages be passed between those devices, and the contents and meaning of the data in those messages must be unambiguous.

The quality of any business or industrial process outcomes depend upon three major foundations:

1. Quality and reliability of hardware (and physical network) supporting it.
2. Quality of design of the process and decision rules. This includes anticipating all contingencies which would influence a decision made independent of human judgment and involvement.
3. Quality of the data at capture, and quality of definition and clarity of data conveyed between devices.

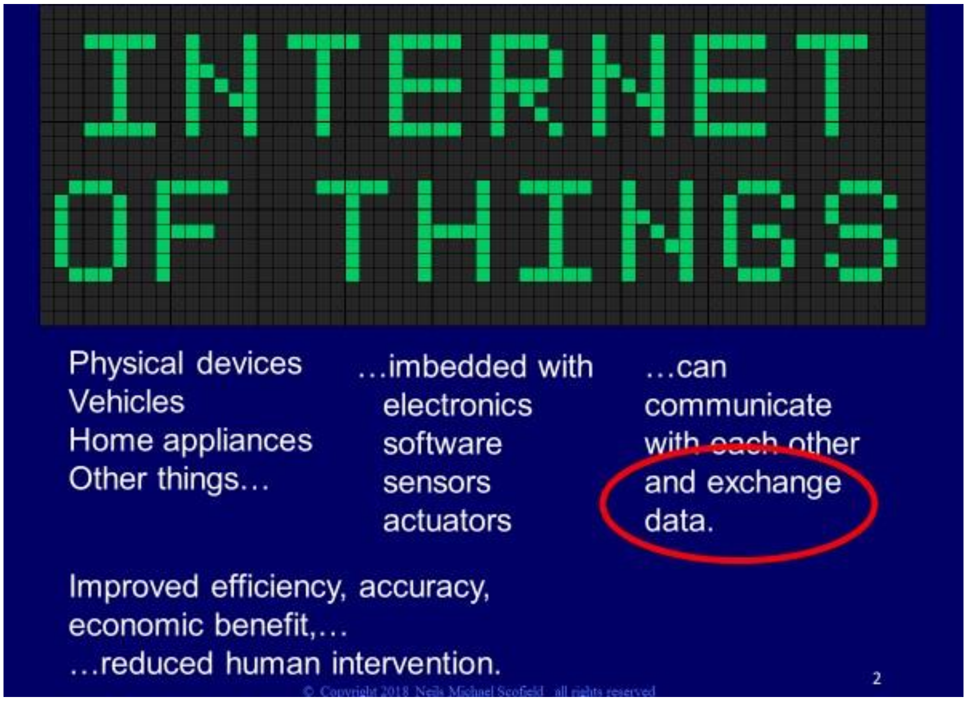
The dilemma is this: for each pair of devices (or sometimes more than just two), who determines the standard of data communication? If a kind of device (like a microwave oven) is supplied by multiple manufacturers, does each have its own design of transaction and the subordinate message structure? If so, what if you have two kinds of microwave ovens in your home?

Each digital message between devices exists in the context of a transaction (conforming to a standard transaction type) which in turn exists in the context of a relationship. The action (or purpose) of a transaction (e.g. start the car engine on a cold morning) may be contingent upon a variety of environmental factors (each observed by some kind of automated sensor). The integration of all that data into a decision can be complex.

How is this all resolved? The answer lies, in part, in a wider understanding of the nature of data, data architecture, and data communication standards, as well as applying the principles of data architecture and metadata to the design of the interactions.

What is the "internet of things"

"Internet of things" is one of the latest buzzwords around which we see a lot of hype (mainly from vendors and manufacturers). What it is essentially is the ability of two devices to share data (or perhaps "information") to achieve some useful domestic or business purposes. That exchange of data generally occurs in some wireless method, although the true internet may or may not be involved.



INTERNET OF THINGS

Physical devices
Vehicles
Home appliances
Other things…

…imbedded with
electronics
software
sensors
actuators

…can
communicate
with each other
and exchange
data.

Improved efficiency, accuracy,
economic benefit,…
…reduced human intervention.

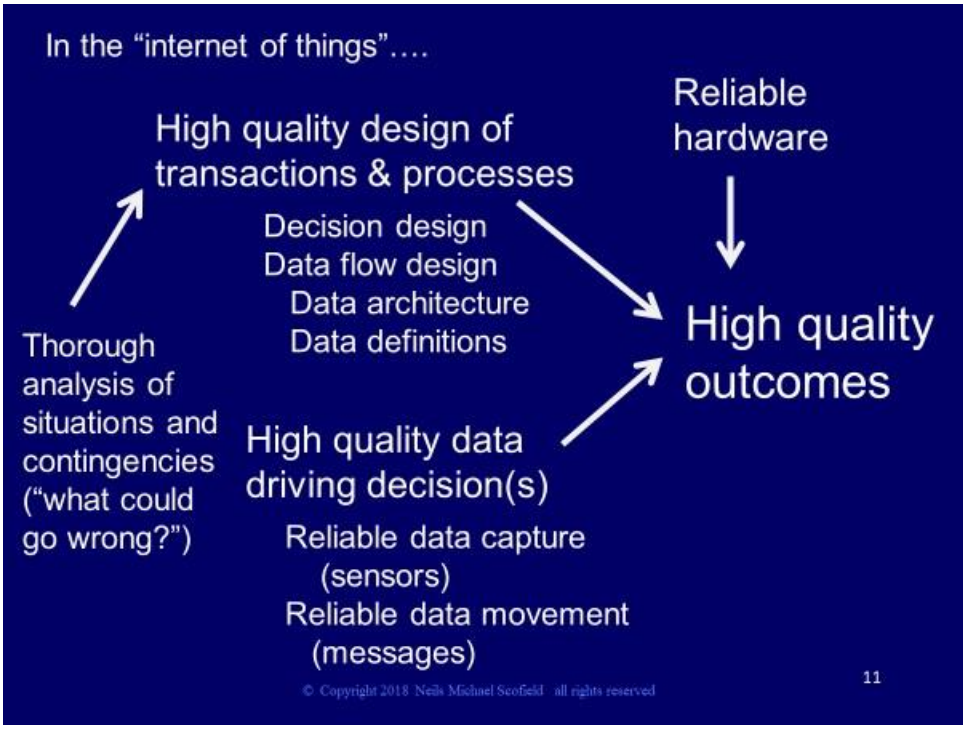
2

© Copyright 2018 Neils Michael Scofield all rights reserved

That part about "reduced human intervention" should raise red flags for anyone interested in quality (or data privacy, for that matter). How do we ensure that the outcome is what is reasonably expected by the stakeholders of the process or exchange?

The "internet of things" is basically where devices can sense and or act, which is not to be confused with the "identification of things" as accomplished in the RFID.

What we must focus upon is the business outcomes of processes which is sought, and its quality.



In the "internet of things"….

High quality design of transactions & processes

Decision design
Data flow design
Data architecture
Data definitions

Reliable hardware

High quality outcomes

Thorough analysis of situations and contingencies ("what could go wrong?")

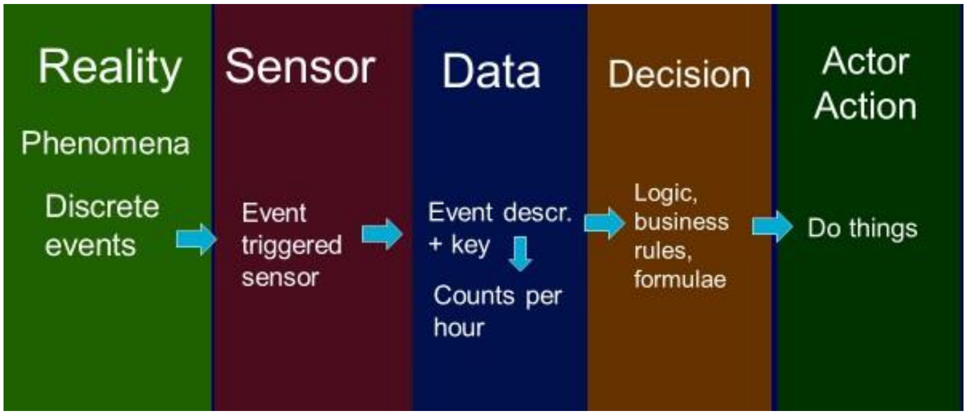
High quality data driving decision(s)

Reliable data capture (sensors)
Reliable data movement (messages)

11

We can organize our thoughts about this topic by recognize five basic elements:

1. Reality
2. Sensors
3. Data
4. Decision
5. Action



Reality
Phenomena
Discrete events

Sensor
Event triggered sensor

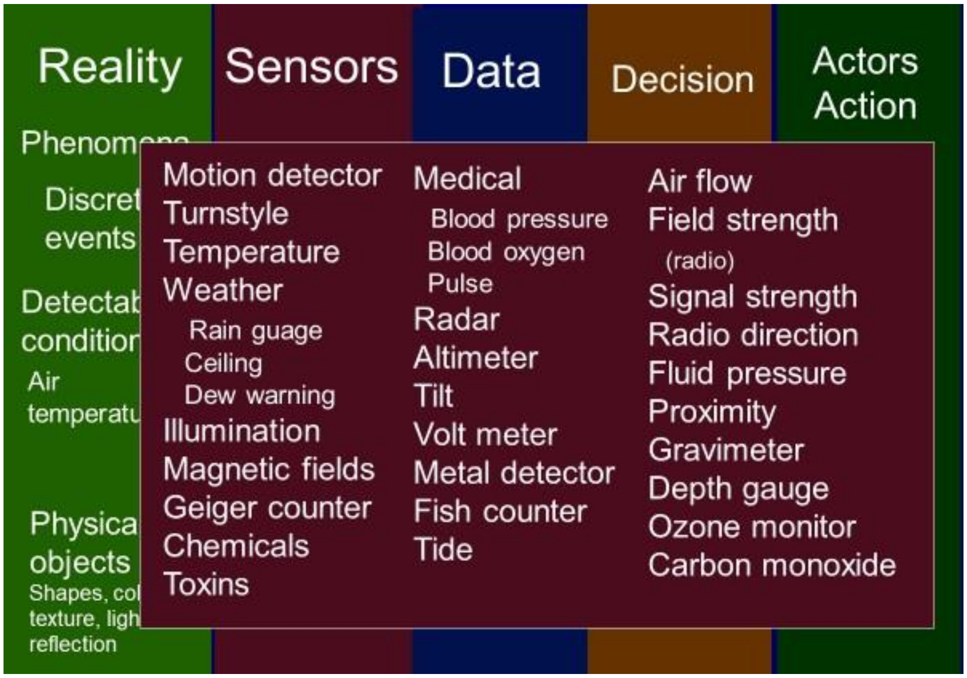
Data
Event descr. + key
Counts per hour

Decision
Logic, business rules, formulae

Actor Action
Do things

Reality (and being in touch with reality) is an essential element for this whole process. The reality can be presence or absences of things or people, or it can be their behavior. This reality (or changes in the reality) must be sensed by sensors.

Sensors create data based on what is happening in reality. A full inventory of all the possible phenomena in the real world, and the sensors to detect them would list over 500 devices. We have a small sample below.



Reality
Sensors
Data
Decision
Actors
Action

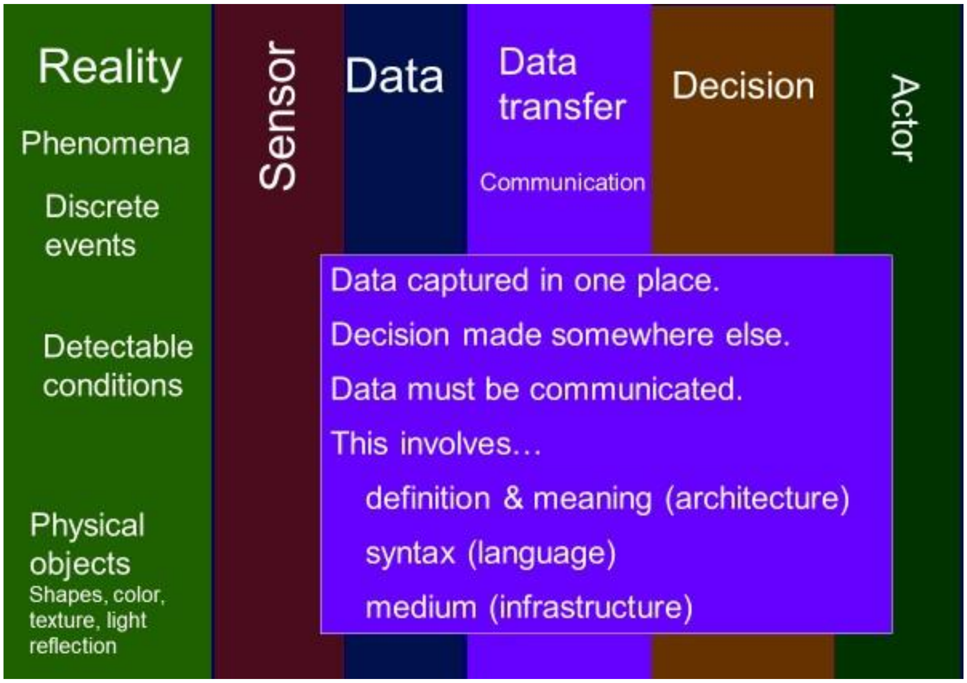
Phenomena
Discret
events
Detectab
conditior
Air
temperatu
Physica
objects
Shapes, col
texture, ligh
reflection

Motion detector
Turnstyle
Temperature
Weather
Rain guage
Ceiling
Dew warning
Illumination
Magnetic fields
Geiger counter
Chemicals
Toxins

Medical
Blood pressure
Blood oxygen
Pulse
Radar
Altimeter
Tilt
Volt meter
Metal detector
Fish counter
Tide

Air flow
Field strength
(radio)
Signal strength
Radio direction
Fluid pressure
Proximity
Gravimeter
Depth gauge
Ozone monitor
Carbon monoxide

Data is what the sensor creates and it is made available to decision-making devices or processes. But there is actually a sixth element to this model: that of **data transfer** or the movement of data. And that involves data quality and data architecture.



Reality
Phenomena
Discrete
events
Detectable
conditions
Physical
objects
Shapes, color,
texture, light
reflection

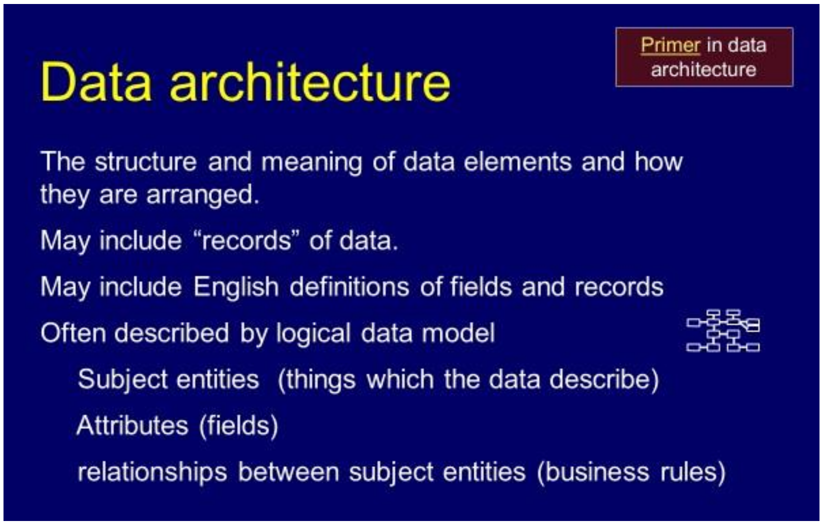
Sensor
Data
Data
transfer
Communication
Decision
Actor

Data captured in one place.
Decision made somewhere else.
Data must be communicated.
This involves…
definition & meaning (architecture)
syntax (language)
medium (infrastructure)

Any time data is moved, a structure (or organization) of the data is required—so that the device (sensor) which creates the data and the device(s) which consume the data (actors or decision-making logic) agree on the meaning and arrangement of the data being passed. Often multiple data elements are involved, and this requires data architecture.

Data architecture

So we must understand the basics of logical data architecture.



Data architecture

Primer in data architecture

The structure and meaning of data elements and how they are arranged.

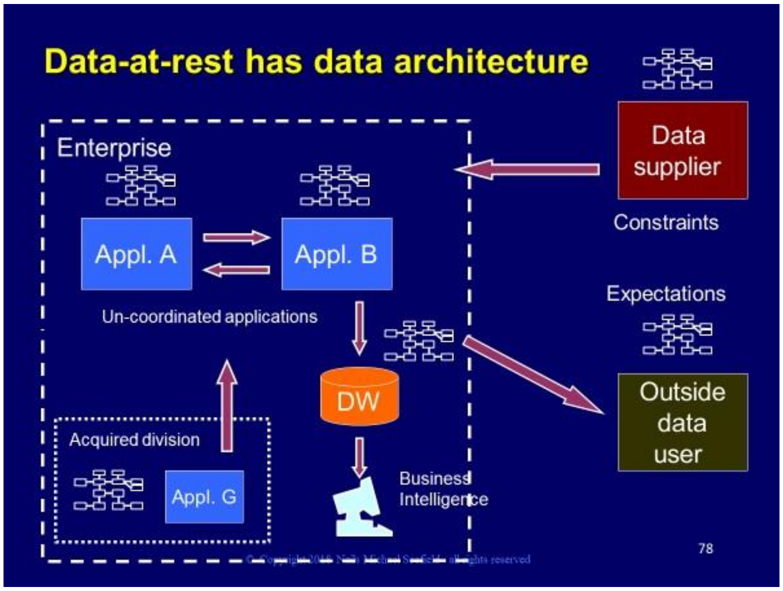
May include "records" of data.

May include English definitions of fields and records

Often described by logical data model

- Subject entities (things which the data describe)
- Attributes (fields)
- relationships between subject entities (business rules)

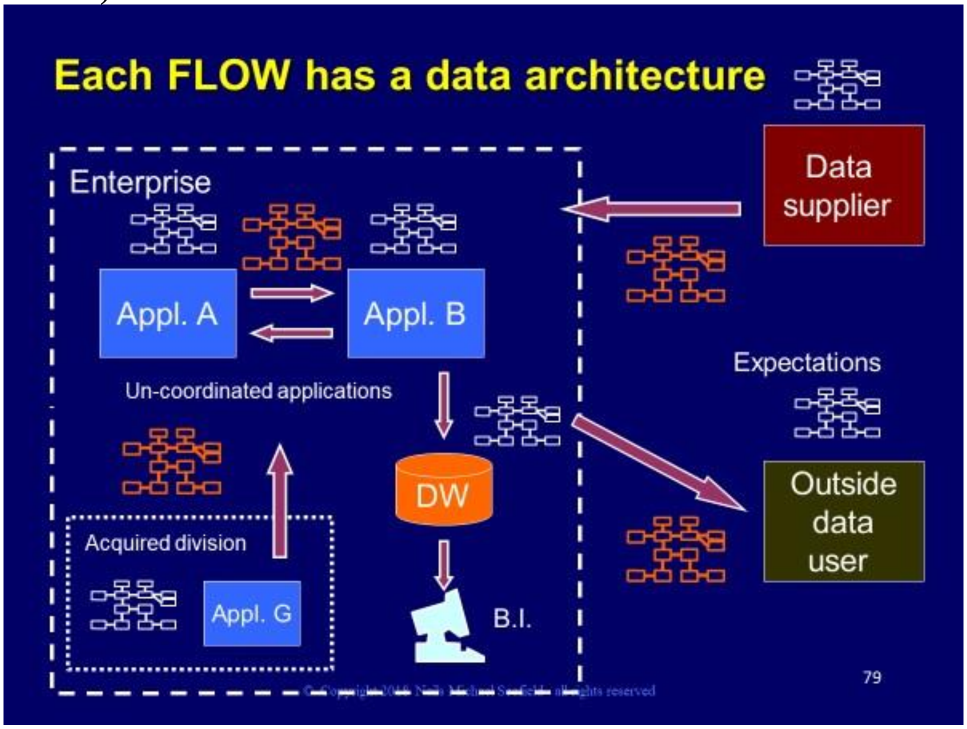
All data "at rest" (in databases or files) has a logical data architecture, whether or not there is a data model to describe it, and whether or not anyone understands it.



Data-at-rest has data architecture

That little white icon (representing an entity relationship diagram) symbolizes architecture for this discussion. Data architecture exists, whether or not a model also exists to describe it.

Similarly, any flow of data (movement from one place to another, or one device to another) has architecture.



Each FLOW has a data architecture

Enterprise

Appl. A

Appl. B

Un-coordinated applications

Acquired division

Appl. G

DW

B.I.

Data supplier

Expectations

Outside data user

© Copyright 2016 Neils Michael Scofield all rights reserved

79

Logical data architecture for most flows of data is rather simple. But it still must be understood by the designers of both ends of the flow—the sensor and the actor or decision-making device.

In most cases, the data flow involves only one simple record. It may or may not have a date-time stamp, and an identification of the origin of the data. But in a situation where the ether may be swamped by multiple messages from multiple sensors, those identification elements become vital.

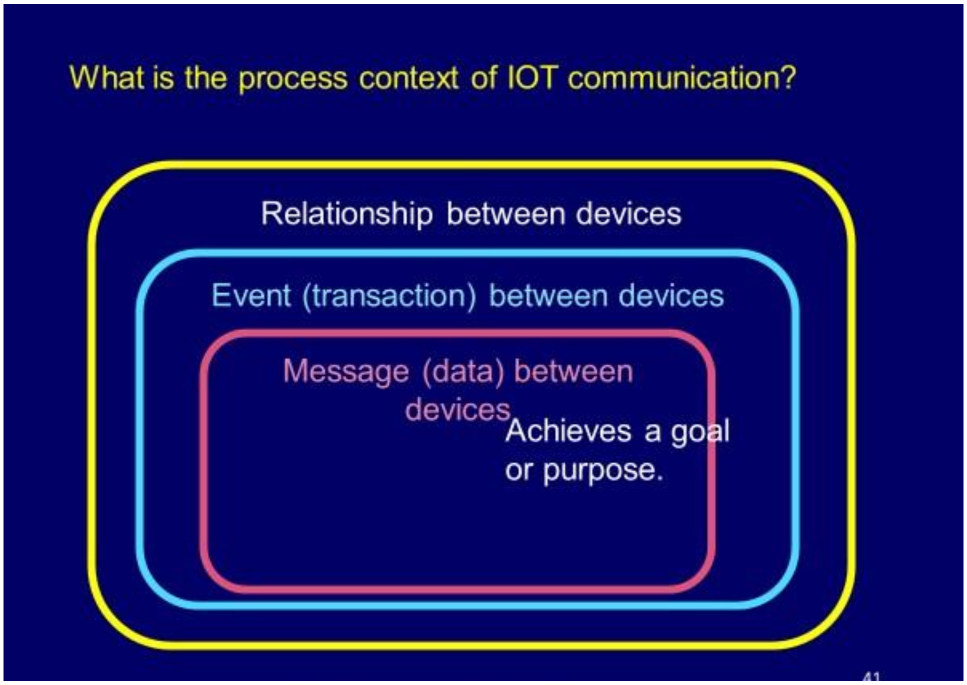
Then, the record may have one or more non-key fields (data elements) which describe what is happening in reality. Examples could be air temperature, humidity, wind direction, and wind speed. But that is a simple record structure.

Some message between devices may contain multiple kinds of records which require more sophistication in the definition of the logical data architecture of the message. For example, a message of patient observations and treatment in a home-health service, may contain one record describing the general situation (time of visit, which nurse visited, patient ID), and multiple secondary (or "child") records of data describing (for example) specific medications administered, or specific treatments rendered.

In such a situation, the transaction consists of multiple kinds of records, and multiple counts of records.

Relationships of devices to accomplish a business purpose

Data flow (which I call "messages") between devices may occur multiple times in a business event.



What is the process context of IOT communication?

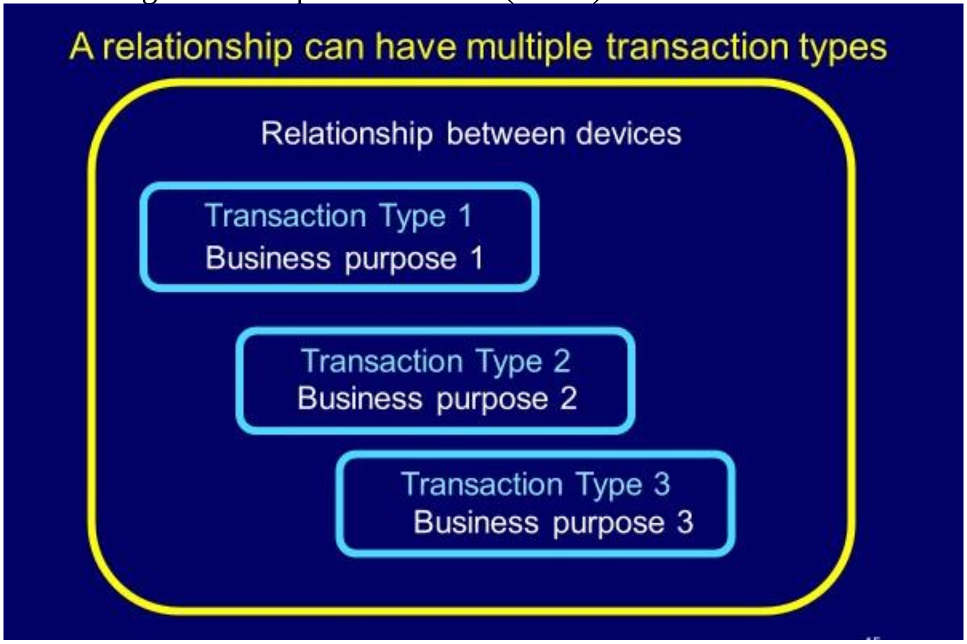
Relationship between devices

Event (transaction) between devices

Message (data) between devices

Achieves a goal or purpose.

And a pair of devices may work together to accomplish several kinds of business events; hence each of those is a distinct "transaction". So subordinate to the relationship between devices might be multiple transactions (below).



A relationship can have multiple transaction types

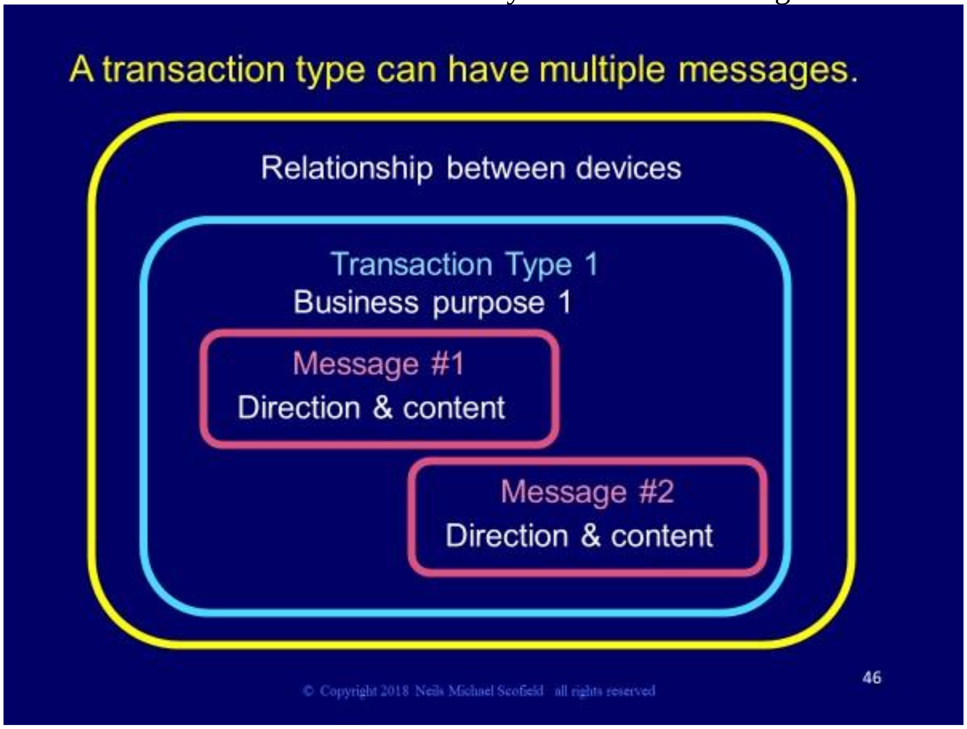
Relationship between devices

Transaction Type 1
Business purpose 1

Transaction Type 2
Business purpose 2

Transaction Type 3
Business purpose 3

And within each transaction may be one more messages.



A transaction type can have multiple messages.

Relationship between devices

Transaction Type 1
Business purpose 1

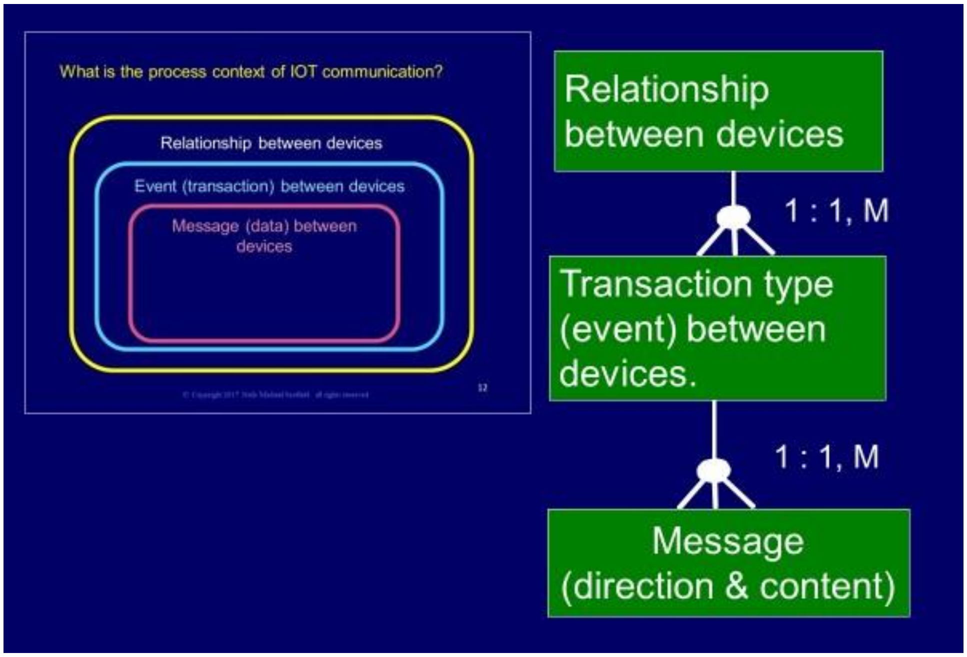
Message #1
Direction & content

Message #2
Direction & content

46

© Copyright 2018 Neils Michael Scofield all rights reserved

Exchanges of multiple messages may occur in a "dialogue" between devices. So each message has a direction (from originating device to recipient) and content (one or more fields of data, highly structured).



What is the process context of IOT communication?

Relationship between devices

Event (transaction) between devices

Message (data) between devices

12

© Copyright 2017 Neils Michael Scofield all rights reserved

Relationship between devices

1 : 1, M

Transaction type (event) between devices.

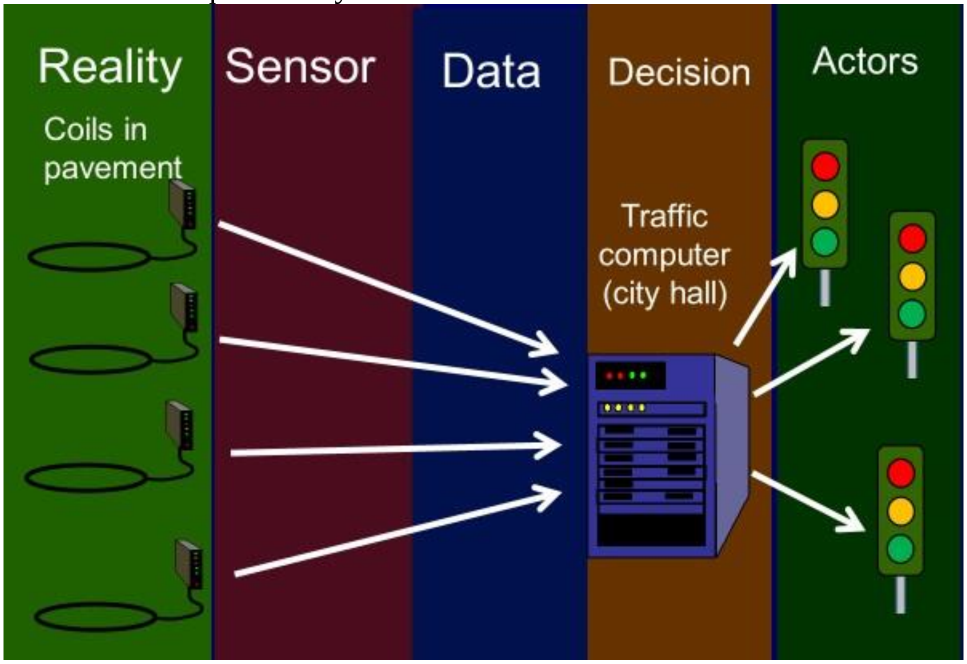
1 : 1, M

Message (direction & content)

We actually can describe this in a simple entity-relationship diagram (green, above).

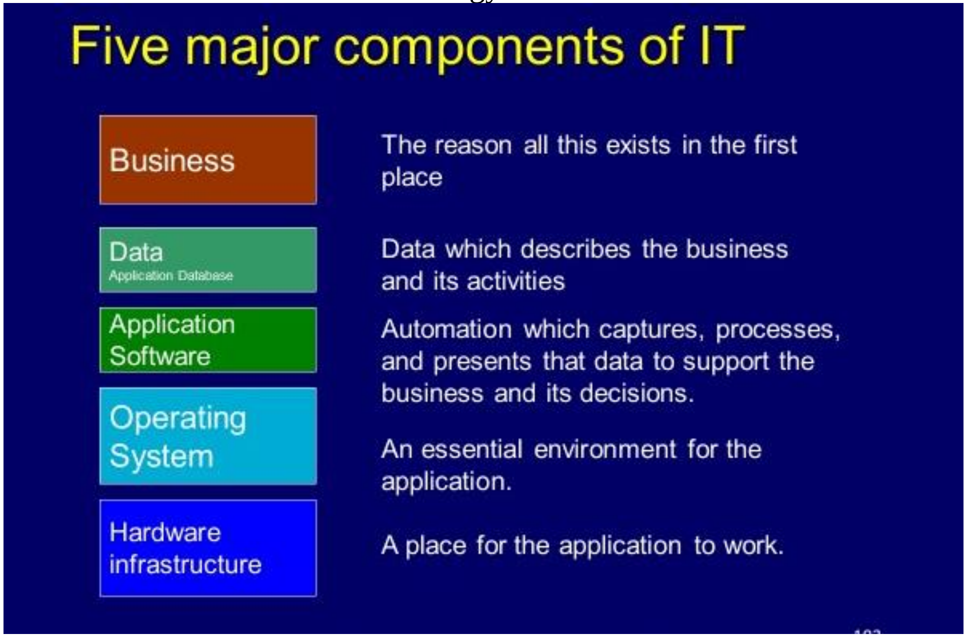
Complex business processes.

Sometimes there are multiple devices involved in a transaction or process. For example, a vehicular traffic control system may involve many sensors, and many traffic signals. They fit into our model quite nicely.



They fit into a network of sensors and actors (traffic signals) with a central, decision-making "hub" (probably a computer in the city hall).

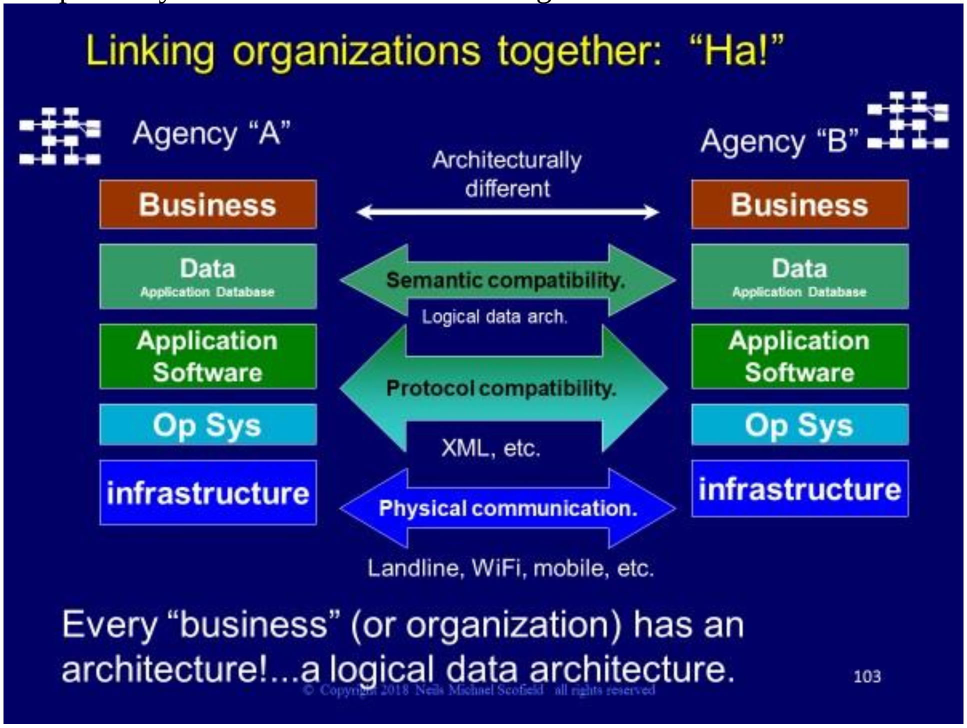
Before discussion the infrastructure of communication between devices, I must review a basic model of information technology.



Five major components of IT

Business	The reason all this exists in the first place
Data Application Database	Data which describes the business and its activities
Application Software	Automation which captures, processes, and presents that data to support the business and its decisions.
Operating System	An essential environment for the application.
Hardware infrastructure	A place for the application to work.

When data is transferred between organizations, there are actually three levels of compatibility to consider about the messages.



Linking organizations together: "Ha!"

Agency "A"

Agency "B"

Architecturally different

Business

Data
Application Database

Application Software

Op Sys

infrastructure

Semantic compatibility.

Logical data arch.

Protocol compatibility.

XML, etc.

Physical communication.

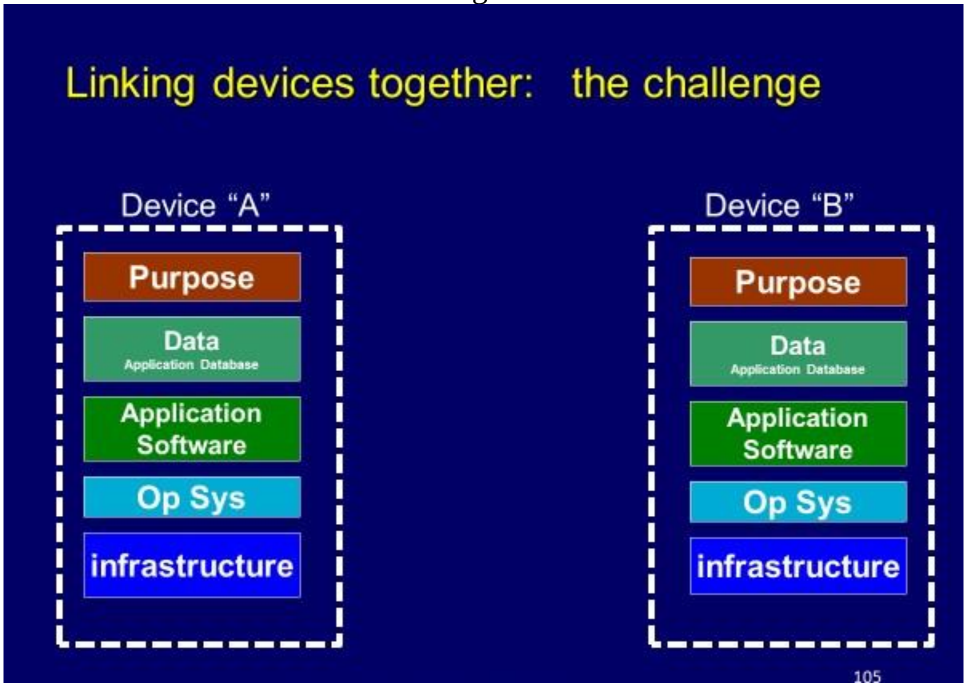
Landline, WiFi, mobile, etc.

Every "business" (or organization) has an architecture!...a logical data architecture.

103

© Copyright 2018 Neils Michael Scofield all rights reserved

The same is true about linking devices.



Linking devices together: the challenge

Device "A"

Device "B"

Purpose

Data
Application Database

Application Software

Op Sys

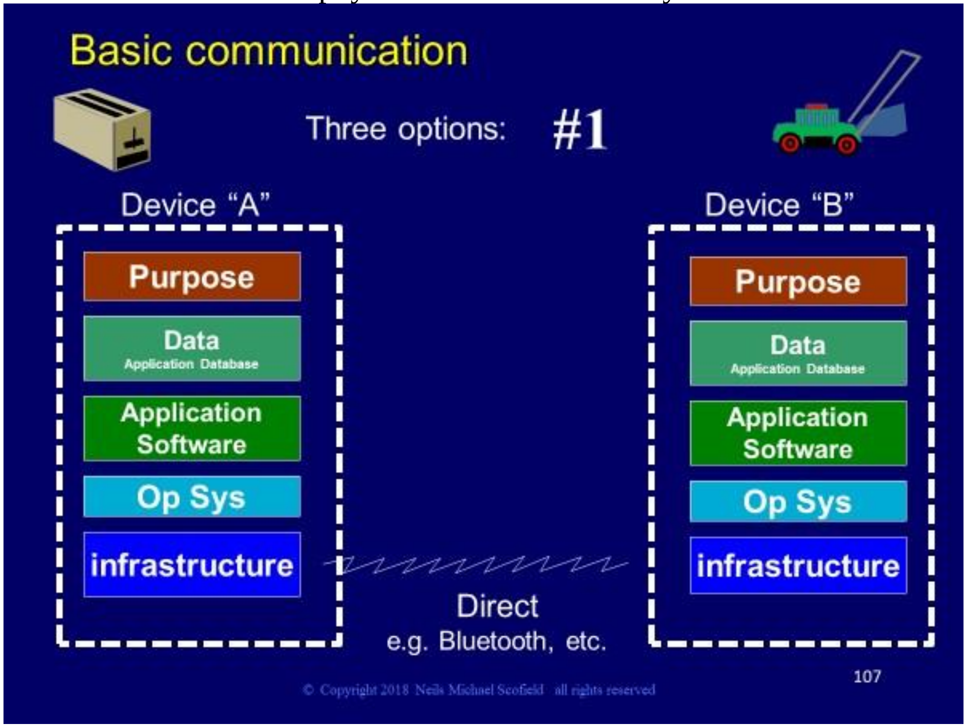
infrastructure

105

Each device has a business purpose, involves data, may have a chip which has both an operating system and application software on it.

Question: How is that software updated if necessary?

There is first the physical communication layer.



Basic communication

Three options: #1

Device "A"

Purpose

Data
Application Database

Application Software

Op Sys

infrastructure

Device "B"

Purpose

Data
Application Database

Application Software

Op Sys

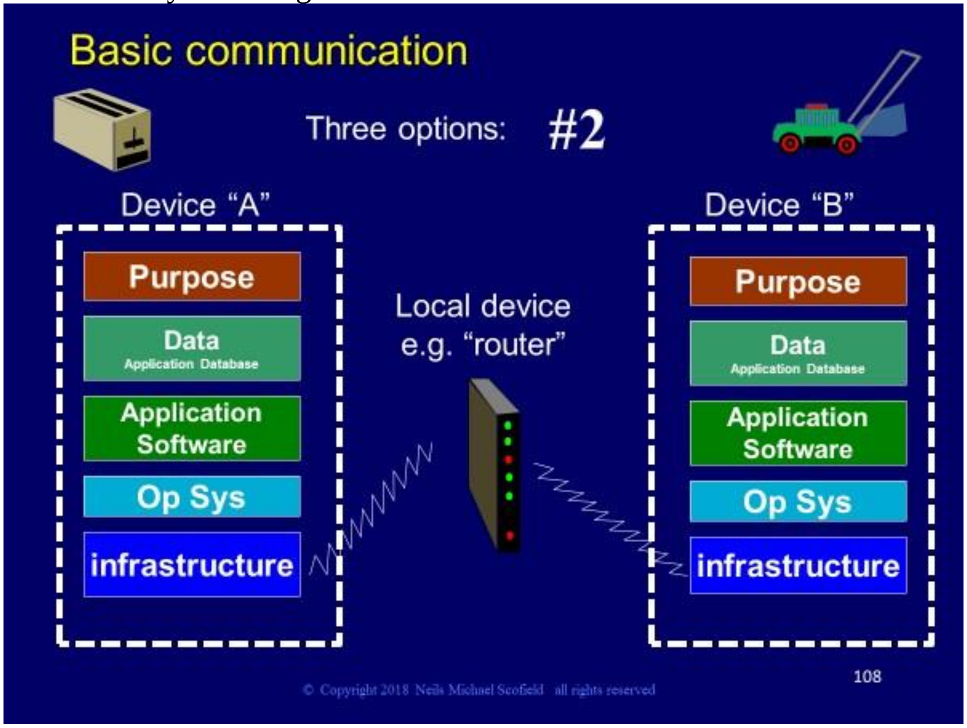
infrastructure

Direct
e.g. Bluetooth, etc.

© Copyright 2018 Neils Michael Scofield all rights reserved

107

It may be through a local wireless router.



Basic communication

Three options: #2

Device "A"

Purpose

Data
Application Database

Application Software

Op Sys

infrastructure

Local device
e.g. "router"

Device "B"

Purpose

Data
Application Database

Application Software

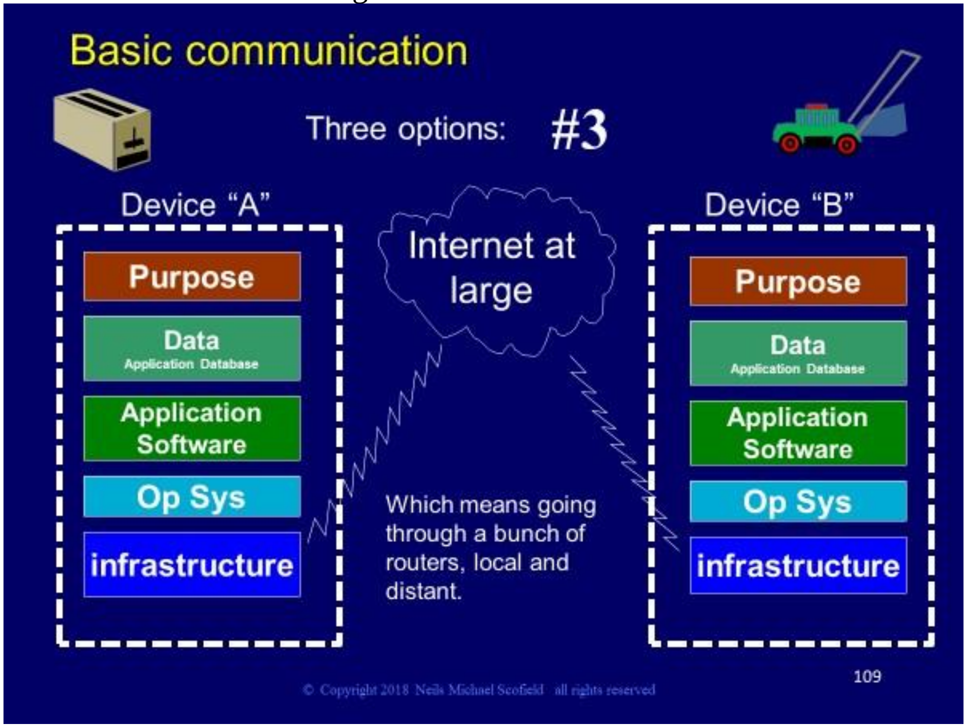
Op Sys

infrastructure

© Copyright 2018 Neils Michael Scofield all rights reserved

108

Or it could be through the internet.



Basic communication

Three options: #3

Device "A"

- Purpose
- Data (Application Database)
- Application Software
- Op Sys
- infrastructure

Internet at large

Which means going through a bunch of routers, local and distant.

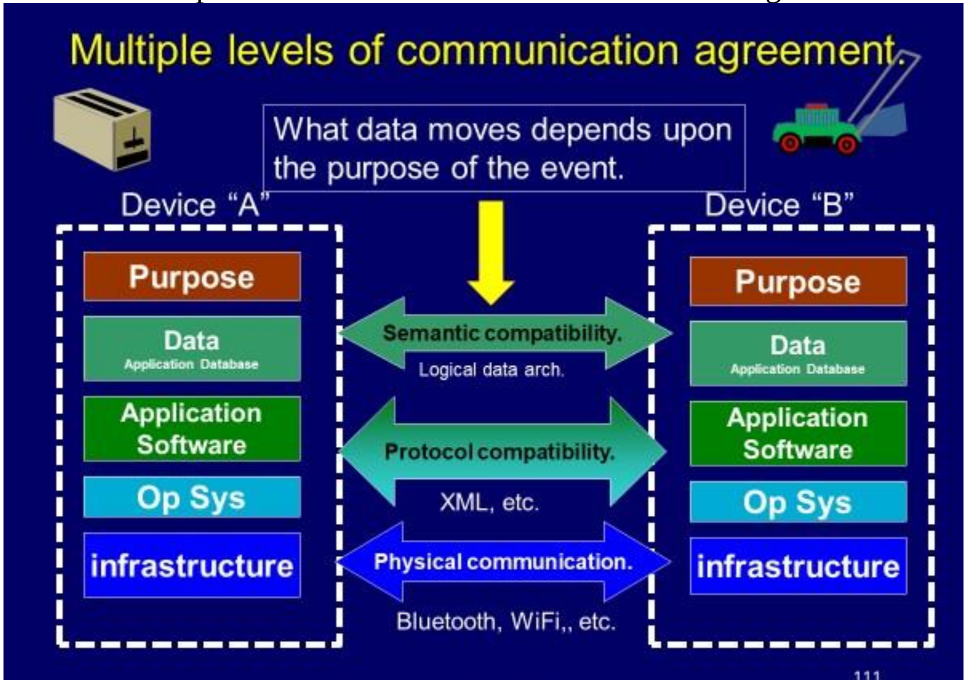
Device "B"

- Purpose
- Data (Application Database)
- Application Software
- Op Sys
- infrastructure

© Copyright 2018 Neils Michael Scofield all rights reserved

109

But all three methods of physical communication are only the physical infrastructure foundation for protocol and semantic structure of the messages.



Multiple levels of communication agreement.

What data moves depends upon the purpose of the event.

Device "A"

- Purpose
- Data (Application Database)
- Application Software
- Op Sys
- infrastructure

Semantic compatibility.
Logical data arch.

Protocol compatibility.
XML, etc.

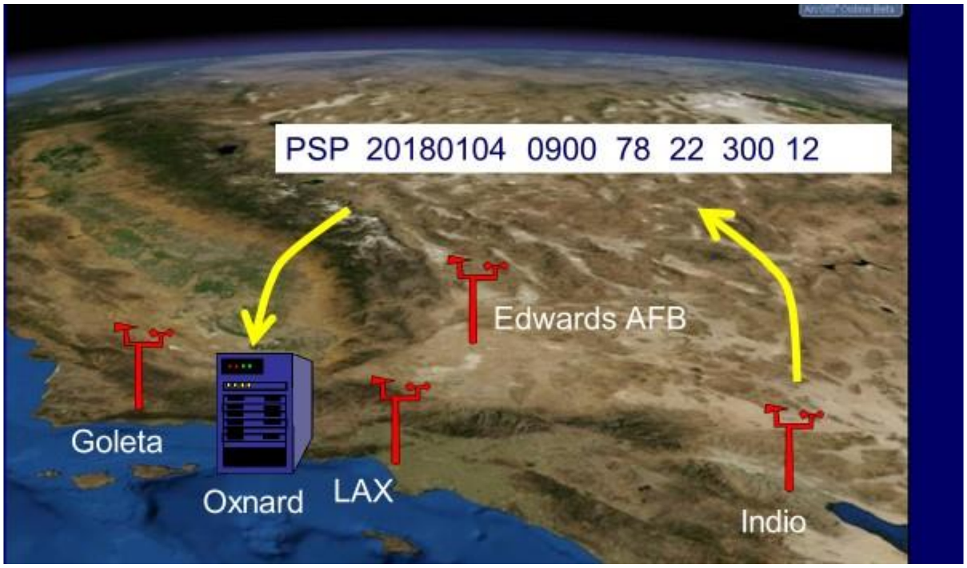
Physical communication.
Bluetooth, WiFi,, etc.

Device "B"

- Purpose
- Data (Application Database)
- Application Software
- Op Sys
- infrastructure

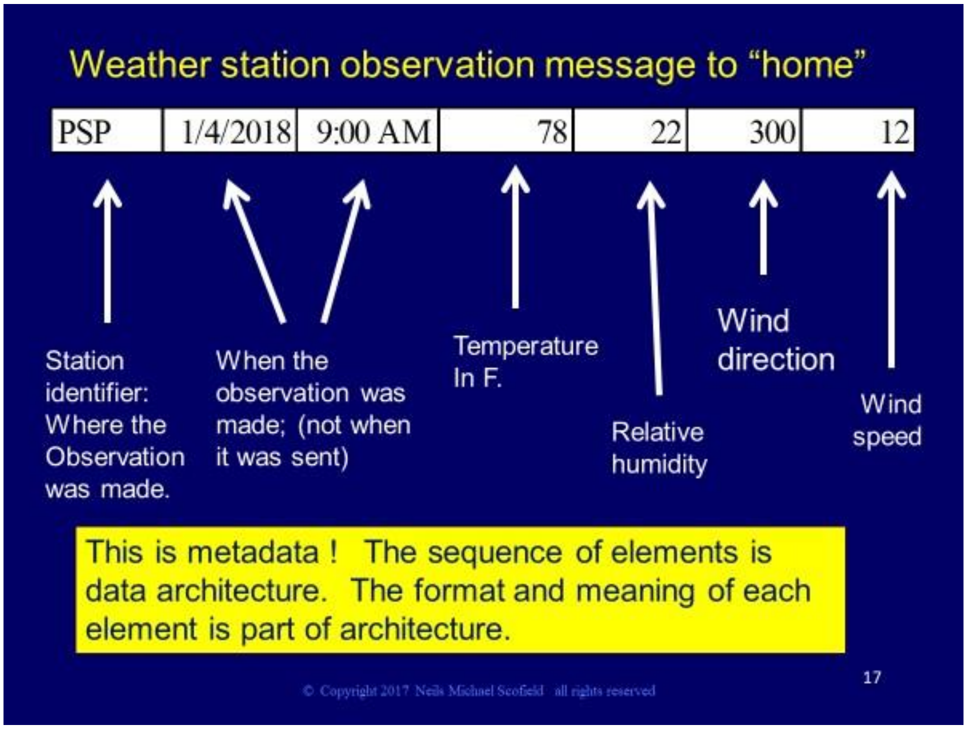
111

The need for well thought-through data record design (data architecture) is evident in how distributed sensors send messages (in this case weather observations) to a central computer. Consider gathering weather data around southern California.



The message has a sequence of data elements, but they are not defined. With well-understood logical data architecture of the message (or record) they don't need to be defined on each transmission.

The record layout which we see above has no delimiters or labels (unless they use XML which is becoming very popular). But the position of each element is agreed upon by the designer of the sensor (sending device) and the designer of the receiving computer.



Weather station observation message to "home"

PSP	1/4/2018	9:00 AM	78	22	300	12

Station identifier: Where the Observation was made.
When the observation was made; (not when it was sent)
Temperature In F.
Relative humidity
Wind direction
Wind speed

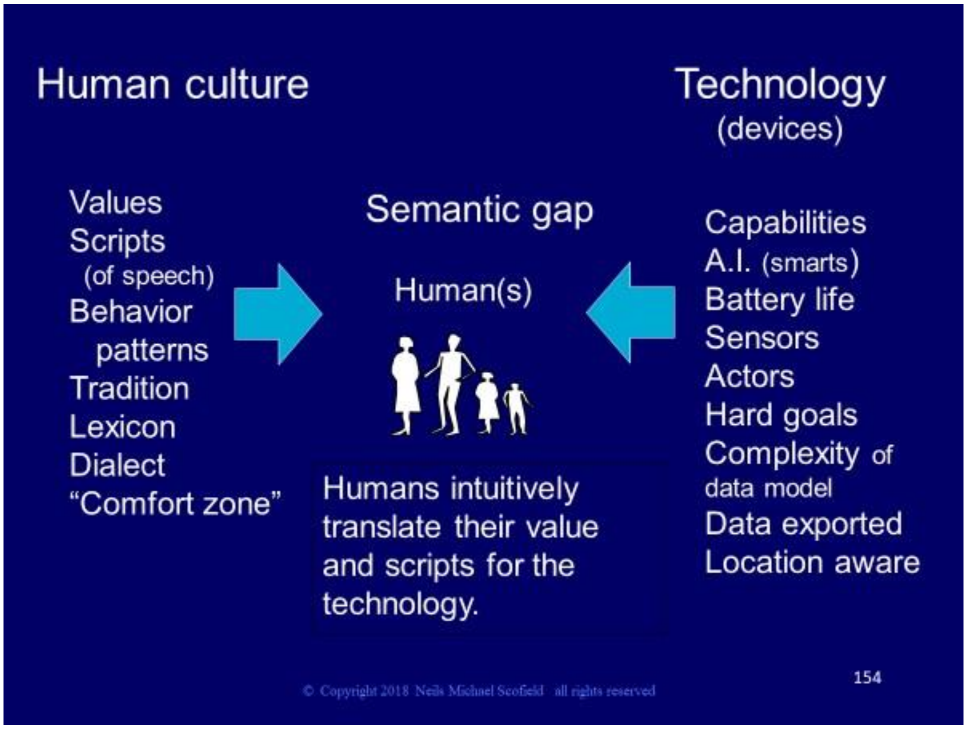
This is metadata ! The sequence of elements is data architecture. The format and meaning of each element is part of architecture.

17

© Copyright 2017 Neils Michael Scofield all rights reserved

This kind of rigor has been around a long time starting with standards for Electronic Data Interchange ("EDI"). But with the potential of more flows (actually more kinds of flows) of data (or messaging) between devices, well-understood architecture and standards are more important.

We must also consider the enormous gap between the rigid logic of devices, and the very open and unstructured nature of human cognition.



Human culture

Values
Scripts
(of speech)
Behavior
patterns
Tradition
Lexicon
Dialect
"Comfort zone"

Semantic gap

Human(s)

Humans intuitively translate their value and scripts for the technology.

Technology
(devices)

Capabilities
A.I. (smarts)
Battery life
Sensors
Actors
Hard goals
Complexity of data model
Data exported
Location aware

© Copyright 2018 Neils Michael Scofield all rights reserved

154

Humans can assess a situation and make a decision, usually intuitively without giving it much thought. This particularly happens when cooking a meal or driving a car. It also happens in criminal detective work. A bright mind can look at a situation, and achieve an "insight" into what is going on that might not have been obvious to the superficial observer.

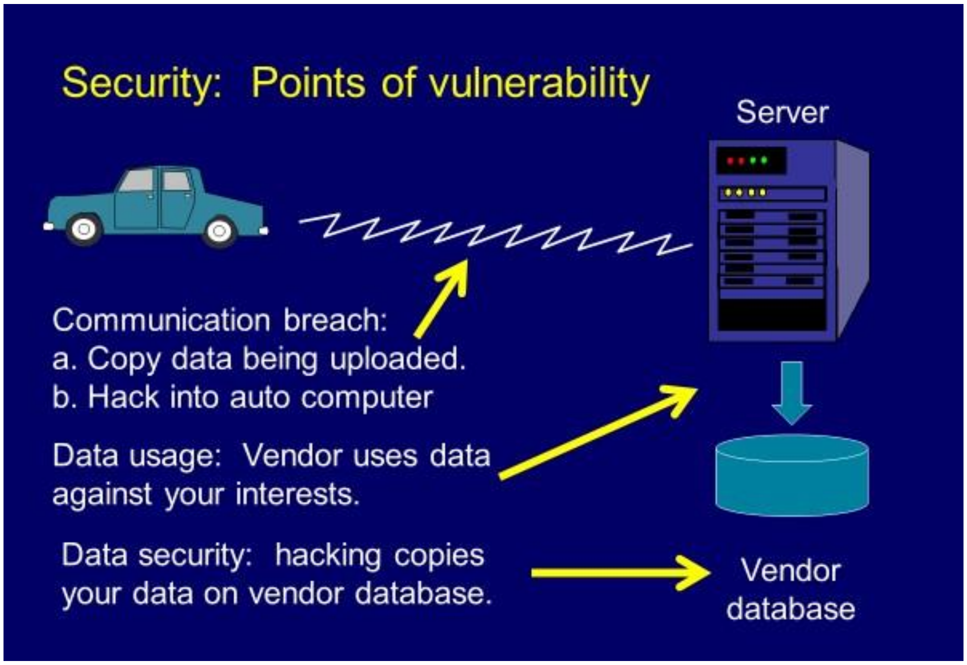
To reduce all the observations and micro-decisions made in the human brain to a set of decision rules embedded in software logic is a daunting challenge.

Issues of data security and safety

Any time personal data is captured and recorded (on paper or electronically) there is potential for that data to be compromised—to be copied without authorization (perhaps for nefarious purposes) or being modified without permission.

We see three general categories of vulnerability:

1. **Transmission of data.** Particularly through wireless means, there is always the potential of intercept and/or altering. True, there are encryption techniques which may reduce the risk, but hacking seems to be a growing art.



Security: Points of vulnerability

Server

Communication breach:
a. Copy data being uploaded.
b. Hack into auto computer

Data usage: Vendor uses data against your interests.

Data security: hacking copies your data on vendor database.

Vendor database

2. **Data usage**. What the host company (vendor or manufacturer) does with the data. They may share your data with "business partners" and you may not be aware of that, or have any say in it.

3. **Data security.** Once the data is sitting on their database (on their premises or in the "cloud") it is still vulnerable to being hacked. There have been many well-publicized security breaches for major companies including Target, Equifax, and others.

These are all considerations when you employ a "smart" device (for personal or business use) which is frequently or constantly "phoning home". You need to thoroughly understand what data is being collected and what is being done with it.

Conclusion

Readers should be somewhat skeptical about the hype behind the "internet of things." We must ask, critically, just what business or functional purpose is sought when we attempt to link two or more devices together so they may interact.

So many questions arise about data security and the quality of outcomes. Any "process of continuous improvement" requires rigorous record-keeping of inputs and outcomes. Where is this data stored? Who stores it? If the devices come from two different vendors, how is that storage accomplished?

And then how does improvement happen? By modification of the design of the process? Who does that?

There are enormous dangers (and some opportunities) in the "internet of things". Consumers and business designers must be very careful.

The End

Michael Scofield, M.B.A. is an Assistant Professor at Loma Linda University, and is a popular speaker in topics of data quality, decision-support, and data visualization to professional audiences all over the United States. He occasionally consults in data quality and data analysis. In addition to his lectures before a variety of professional organizations (including DAMA chapters, TDWI chapters, American Society for Quality chapters, Institute of Internal Auditor chapters, etc.) he is also a frequent guest lecturer at a number of universities.