C4R

Capacity for Rail

Ubiquitous Data for Railway Operations

Digital Operations for Enhancing Performance and Capacity Workshop

Olomouc, 27 – 28 April 2017

John Easton

Work Package 3.4 Leader

UNIVERSITY OF BIRMINGHAM

SEVENTH FRAMEWORK PROGRAMME

Work Package 3.4 – Aims & Objectives

Capacity for Rail

Develop a data architecture that is able to provide ubiquitous data for railway operations and supporting applications

- Understand the data exchange and integration requirements of railway operations;
- Provide extensions to existing data notations that support operational data;
- Develop new data model supporting autonomous data exchange and reasoning;
- Develop appropriate architectural frameworks for distributed processing in railway operations.

Breakdown of Work

- Task 3.4.1 - Review of existing processes, practices and tools for operational data management (M6)
- Task 3.4.2 - Data notation development (M21)
- Task 3.4.3 - Autonomous data exchange (M21)
- Task 3.4.4 - Data architecture development (M24)
- Task 3.4.5 - Proof of concept for data modelling and architecture (M33)

- D3.4.1 – Data notation and modelling
- D3.4.2 – Data architecture

Data Modelling State of the Art

Review of existing models, standards and open data curation / provisioning w.r.t. the requirements of the storyboards

- Supports targeted development activity for demonstrations:
  - Current “best practice” e.g. XML;
  - Next generation models e.g. Ontology;
  - Linked Open Data.
- Inform development and demonstration exercises

Storyboard 1

Capacity for Rail

Storyboard 1: Single source, multi-scaled infrastructure data for operations and simulation.

2020

Multiple data resources for the same infrastructure contain potentially conflicting views of the same physical assets. Data is presented using a range of models, limiting the easy integration of data from multiple IMs; as a result manual alignment and integration of the data is required before cross-border simulation or planning work can be performed.

2030

Rail infrastructure data is provided in a common, open format by IMs at multiple, pre-defined levels of granularity appropriate to operational and planning tasks.
Automated consistency checks ensure old or conflicting data is flagged for review by expert staff.

2050

Distributed, single sources of truth for rail infra. data are available at the IM level, thanks to common concept-based treatment of models. Detailed data can be auto-aggregated to support planning, operations and simulation of the infrastructure at any granularity (micro, macro etc.). Data models relate to over-arching representations of the railway system, allowing interoperability of data across domains.

Storyboard 2

Storyboard 2: Data integration as a driver for effective usage of cross-mode capacity; long distances routes are fed more efficiently, and handling of disruption is improved.

2050

Good utilisation of capacity allows for frequent services. Flows of passengers can be controlled at peak times via dynamic modal advice to individuals, managing their arrival / departure at long distance terminals and ensuring they feed into the system via all available local services. Disruptions are handled seamlessly, supported by e-tickets that allow free-flow of passengers between available transport modes.

2030

Passengers have easy access to data on alternative connections, inc. those that run via different modes.
Open ticketing information allows informed decision-making, allowing the correct trade-off between cost, speed of transit, and available facilities to be made.

2020

Disruption results in lengthy delays for passengers as they wait for resolution of issues / later services. Passengers and staff may be left out-of-position for follow-on travel.
Transport modes that offer well-known alternative connections may be heavily congested while less obvious, but no less valid route choices have spare capacity available.

Storyboard 3

Storyboard 3: Real-time data in support of cross-border and cross-organisation operations.

2050

Live views of network and asset state available at high resolution to operational staff. Real-time service and routing information available to all involved railway undertakings through concept-based models for infrastructure, operational practices, and asset condition. Real-time network information allows more accurate prediction of service arrivals and optimisation of live traffic plans.

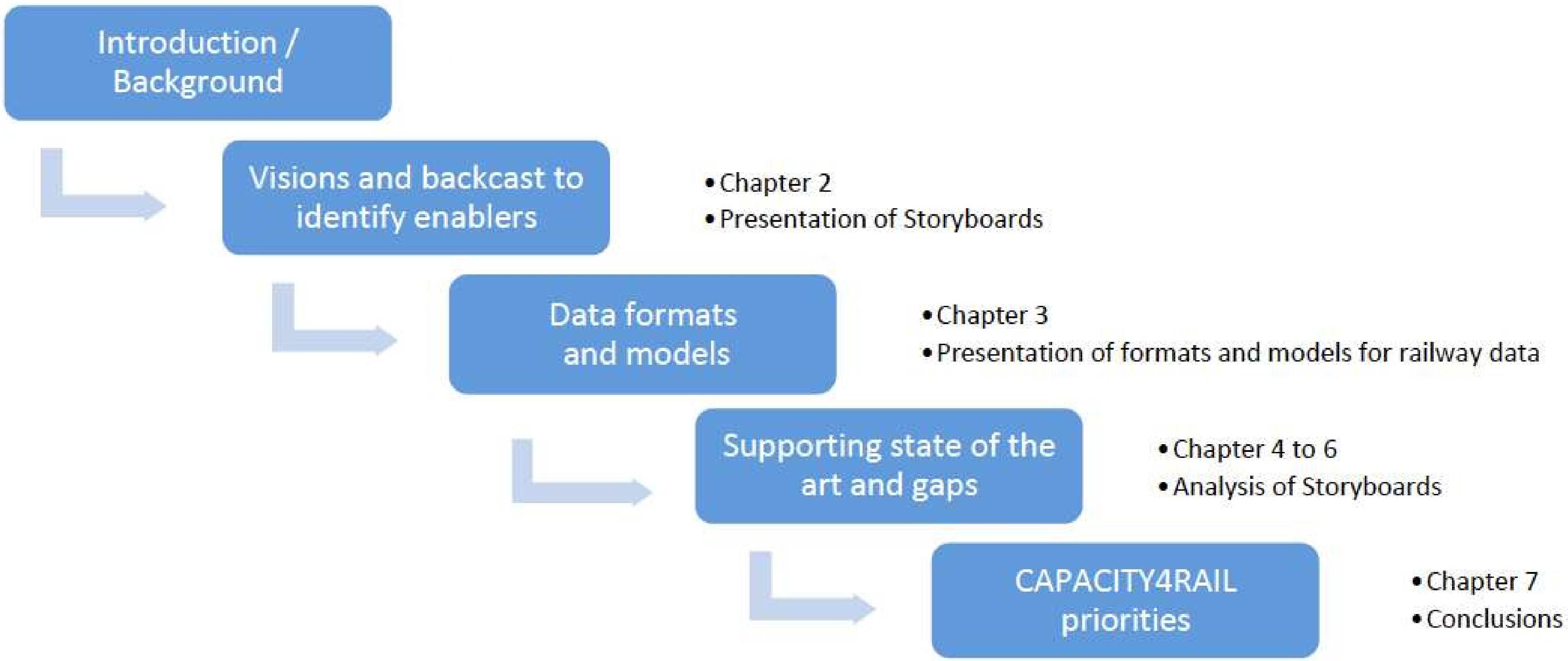
2030

Cross-border service data available at appropriate levels of granularity and in consistent models for all operators / undertakings along the planned route. Changes to schedule of services pushed in real-time to undertakings as updated traffic plans.

2020

Plans for cross-border services available to undertakings involved, but data is at varying levels of granularity and not necessarily up to date if on-route changes have been made or disruptions are involved.

Structure of D3.4.1



Models Reviewed

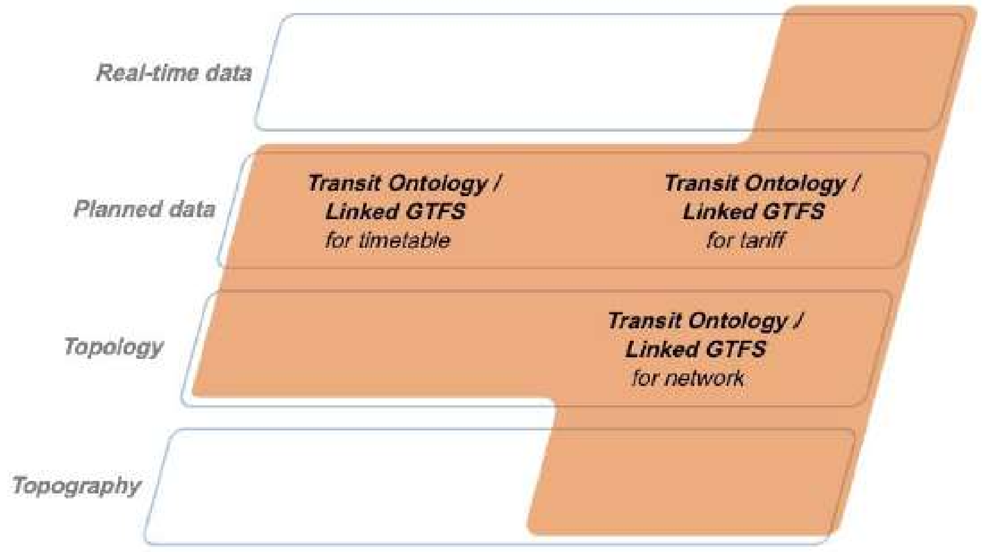
Capacity for Rail

XML-based Models	Non-XML Models	Emerging Approaches
railML 2	TAP TSI	Rail Core Ontology (RaCoOn)
RailTopoModel / railML 3	Google Transit (GTFS) and Real-time	Railway Infrastructure Ontology (RI*)
Register of Infrastructure (RINF)	OSM / ORM	Enriched GTFS (Transit ontology)
Infrastructure for Spatial Info. in the EC (INSPIRE)		Linked Open Data (NEPTUNE)
IDM^{VU}		Public Transport Ontology of Keller, Brunk, & Schlegel
Network and Timetable Exchange (NeTEx)		Semantic Sensor Network (SSN)
Service Interface for Real-time Information (SIRI)		
TAF TSI		
ON-TIME (RTTP)		

Analysis of Models w.r.t. Requirements

Data format / Data Granularity	railML 3	IDM^{VU}	INSPIRE	RINF	OSM
Corridor	Possible	Out of scope	Not available	Out of scope	Out of scope
Macroscopic	Possible	Possible	Possible	Possible	Possible
Mesoscopic	Possible	Out of scope	Possible	Out of scope	Possible
Microscopic	Possible	Possible	Out of scope	Possible	Possible

Storyboard 1: Infrastructure data for operation and simulation



Storyboard 2: Effective usage of cross-mode capacity

Deliverable D3.4.1

Capacity for Rail

Summarised in D3.4.1 - “Data Notation and Modelling”

- State-of-the-art in data formats, models and concepts:
  - railML, RailTopoModel (IRS 30100), RINF, INSPIRE etc.
  - GTFS, OSM, RaCoOn, semantic sensor data etc.
- SB1 – Cross-industry infrastructure data
  - Single, multi-scaled source of infrastructure data that is consistent at microscopic, macroscopic, etc. levels
- SB2 – Effective usage of cross-mode capacity
  - Information resource link between rail and multimodal system, bringing together information from all systems to support operations
- SB3 – Real-time operational data
  - Provision of diverse, high-velocity data resources in support of railway operations (SP4 link)

SEVENTH FRAMEWORK PROGRAMME

Take-home Message

Capacity for Rail

"A key element of success will lie in handling the relationship between rail and other modes…"

Concept	Model
Network topology	RailTopoModel / railML 3
Topology (fallback / degraded)	Open Railway Map
Fixed asset configuration data	railML 3
Sensor data (asset status etc.)	L.O.D. format – SSN or similar
P.I.S. (paths, fare models, ticketing)	NeTEx
Timetables	railML (rail), NeTEx (multimodal)

- All data concepts would benefit from ontology representations, enabling use as Linked Open Data resources

Recommendations (D3.4.1)

- Suggested topics for further development:
  - Interaction of “formal” IM and community-sourced OSM datasets
  - Development of ON-TIME RTTP to align with RailTopoModel requirements
  - Interaction of sensor data (SP4) and RTTP in support of operations
  - Approaches to railway data validation / verification
  - Development of ontology representations for concepts in key XML formats associated with multimodal travel (railML, NeTEx etc.)
  - Demonstration of ontology in support of WP3.4 storyboards

Structure of D3.4.2

Capacity for Rail

Introduction / Background

- Introductions
- Background material linking to D3.4.1

ICT Architectures for Rail Operations

- Section 2
- ESB architectures, InteGRail, ON-TIME, and recs.

Semantic Data Models and Associated Architecures

- Chapter 3
- The Rail Core Ontology (RaCoOn) and data management

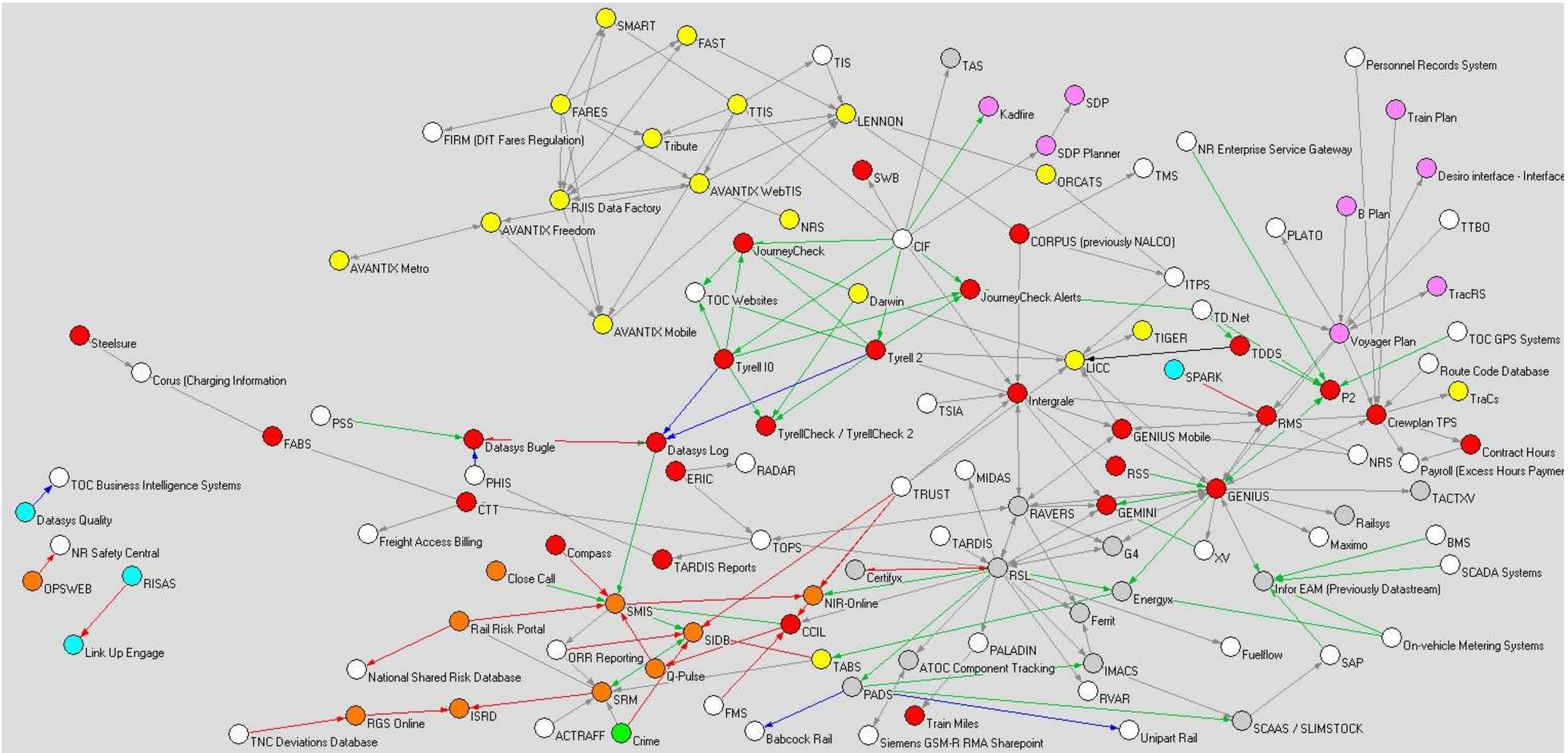
Open and Crowdsourced Data for Improved Situational Awareness

- Chapter 4
- Open data supporting operations

Conclusions

- Chapter 5
- Conclusions

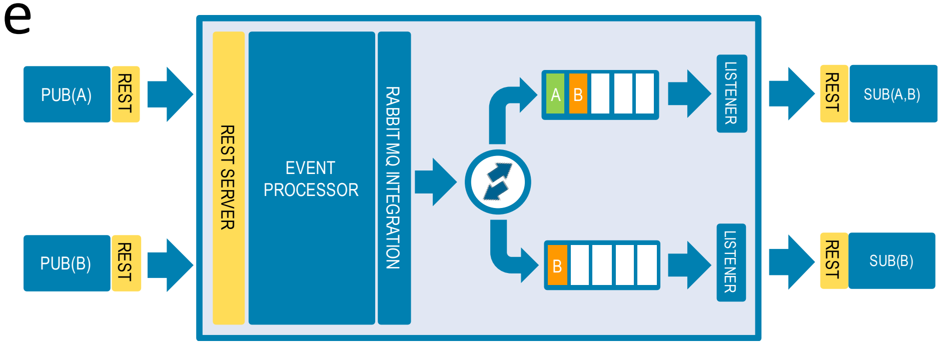
The Complexity of Software in Rail

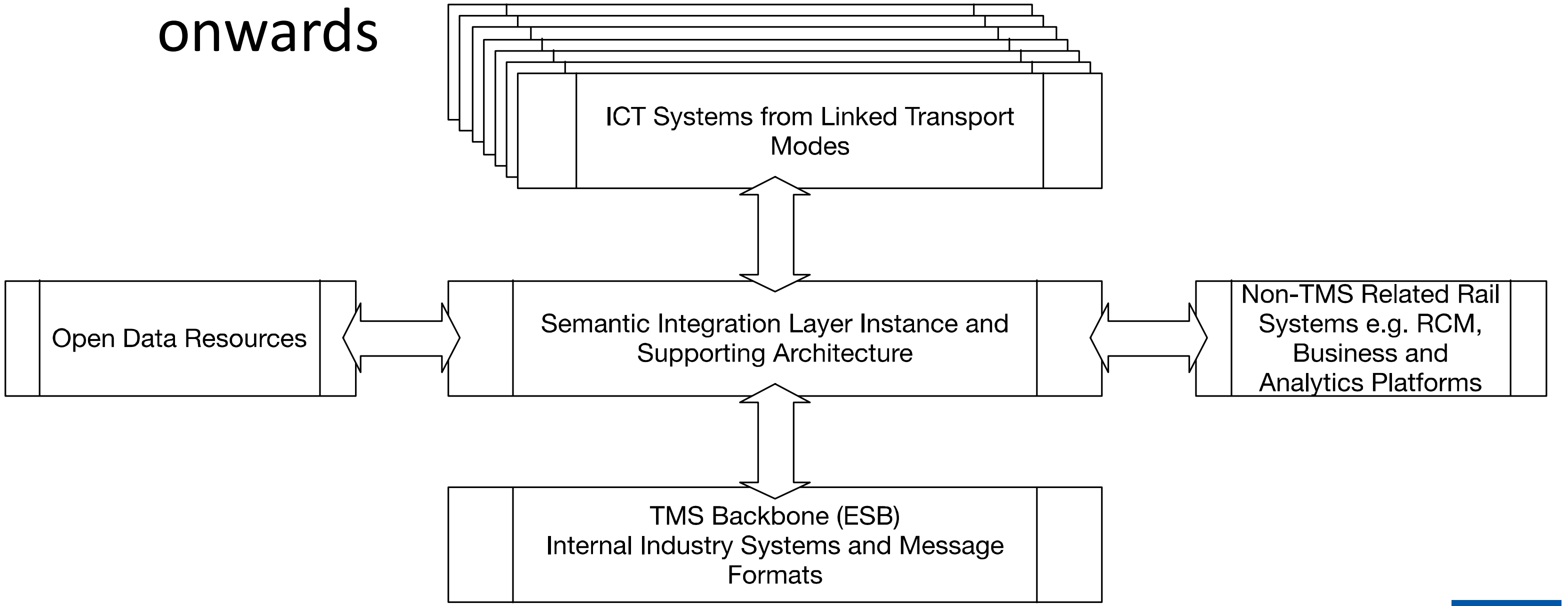


ESB Architecture for Conventional Systems

Best practice (ICT) is applicable

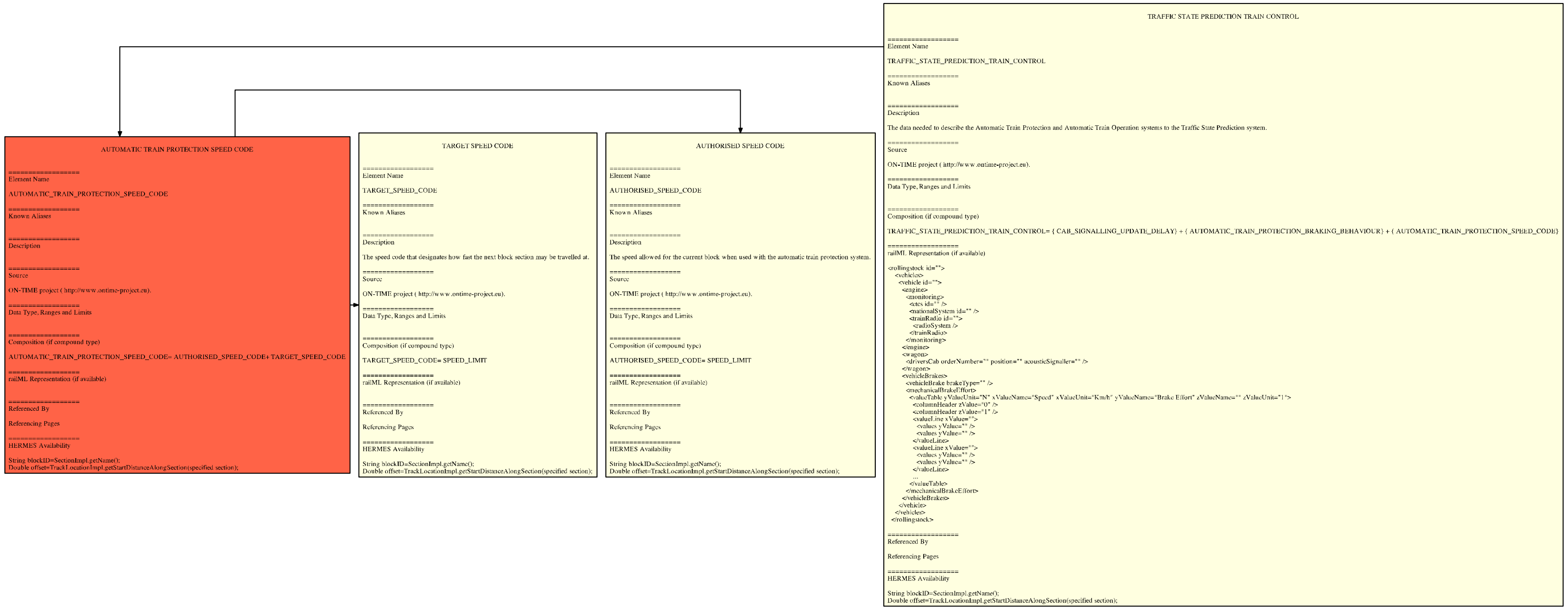
- Adopt recommendations from ON-TIME and others!
  - Builds on a decade of work from InteGRail onwards





Extensions of ON-TIME Dictionary

- In conjunction with Network Rail
- ON-TIME data dictionary as a basis
- Inclusion of concepts around multimodal transport & ETCS
- Extraction of nearest neighbour graphs for each concept



Ubiquitous & Autonomous Data

Capacity for Rail

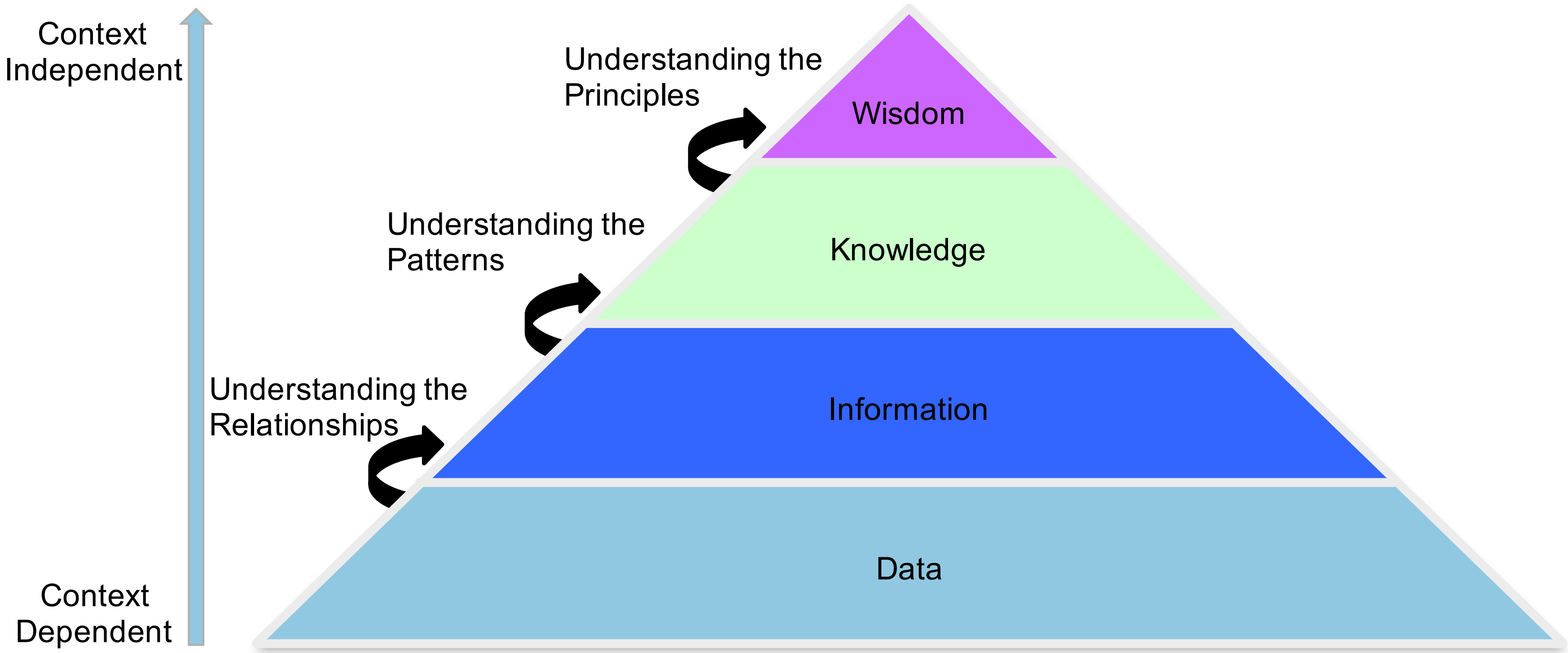
- Globally, we generate huge volumes and varieties of data at increasingly high speeds
  - Data has value in supporting travellers and railway operations
- Resources are increasingly dynamic
  - Sensors may be installed for short periods etc.
  - Resources may become unavailable for reasons outside the user's control
- Challenge
  - Leverage data to gain added value
  - Safe fallback positions when data is unavailable

SEVENTH FRAMEWORK PROGRAMME

Linked Open Data

- Data enrichment by linking to open data provided by other parties
  - Linking a station identified in OSM data, to vehicle movements from NR, passenger schedules provided by ATOC, and building history from DBpedia
- Up-to-date data is accessible via non-volatile URIs
- Decoupled data architecture where single sources of truth are responsible for content
  - Improved accuracy & consistency across services

A Matter of Context

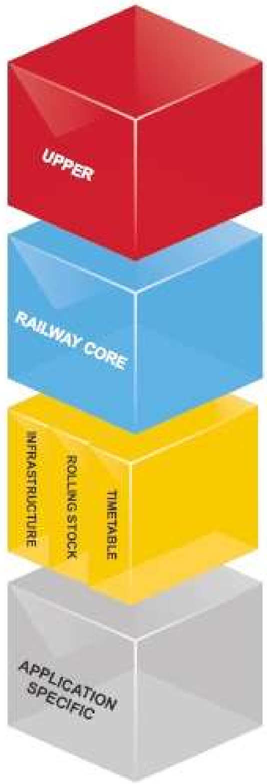


Semantic Web Data Models

- Data integration across system boundaries is a reality of modern ICT architectures, but…
  - …separating data from its original context leads to context and meaning being lost
- Ontology allows the “meaning” of the data to be maintained outside its original system
  - Data is marked-up using tags that reference a published, extensible model of the world-space
  - Public models inherit principles from W3C and other models allowing different views of the same data (e.g. models from different industries) to be used in combination
  - Retention of context means that automated reasoning on data and rules in the model becomes possible
  - Increasing usage in other similar domains (Oil & Gas, Steel)
  - Rail data projects (InteGRail, RRUK Factor20, ON-TIME, C4R)

RaCoOn

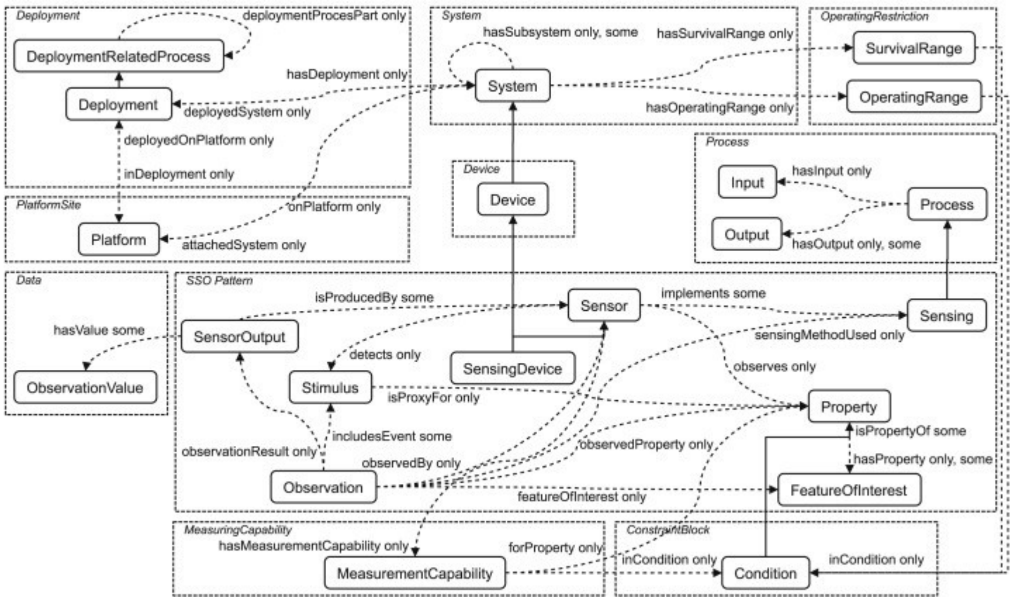
Railway Core Ontology (RaCoOn):

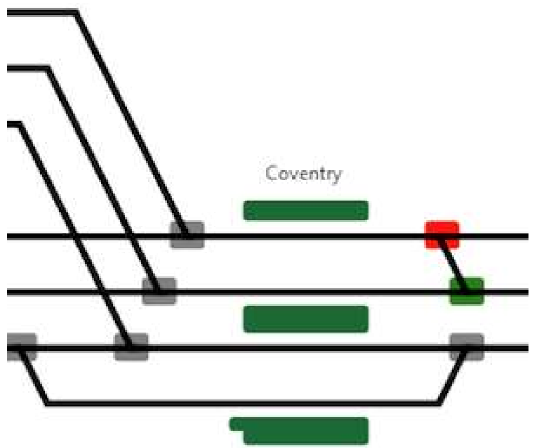


- Support semantic web applications for rail;
- Integrate with external models for multimode:
  - Transit ontology (timetables, networks);
  - Linked NaPTAN (access points);
  - Public transport services ontology.
- Bring in data from other domains to support operations (link to SP4):
  - Semantic Sensor Network, SSN (sensor / RCM data);
  - W3C provenance, PROV-O (understanding of data origins).

SSN

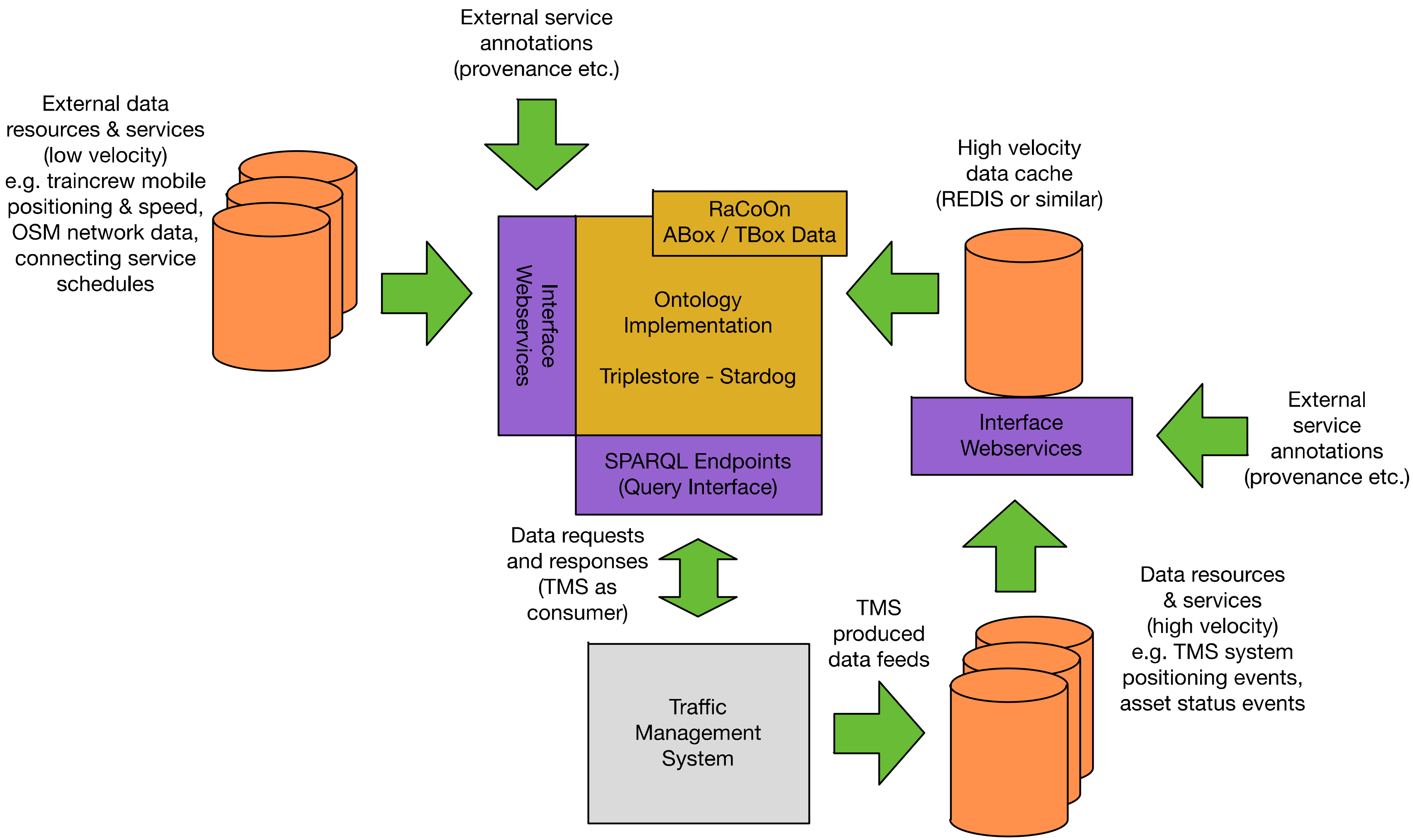
- Developed by W3C
- Describes observations and sensors
- Enables users to work at abstract level
- Combined with event model and LOD such as geodata, allows sensor data to be placed in context of transport system





M. Compton et al. (2012) “The SSN ontology of the W3C semantic sensor network incubator group” Journal of Web Semantics, Volume 17, December 2012, Pages 25-32

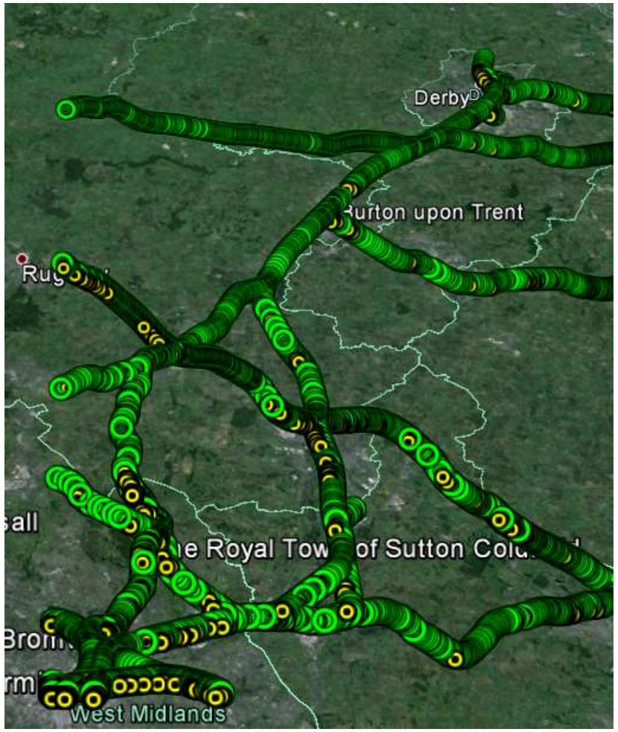
Architecture for Ubiquitous Data



Improved Situational Awareness

Extra information during disruption using crowdsourced data

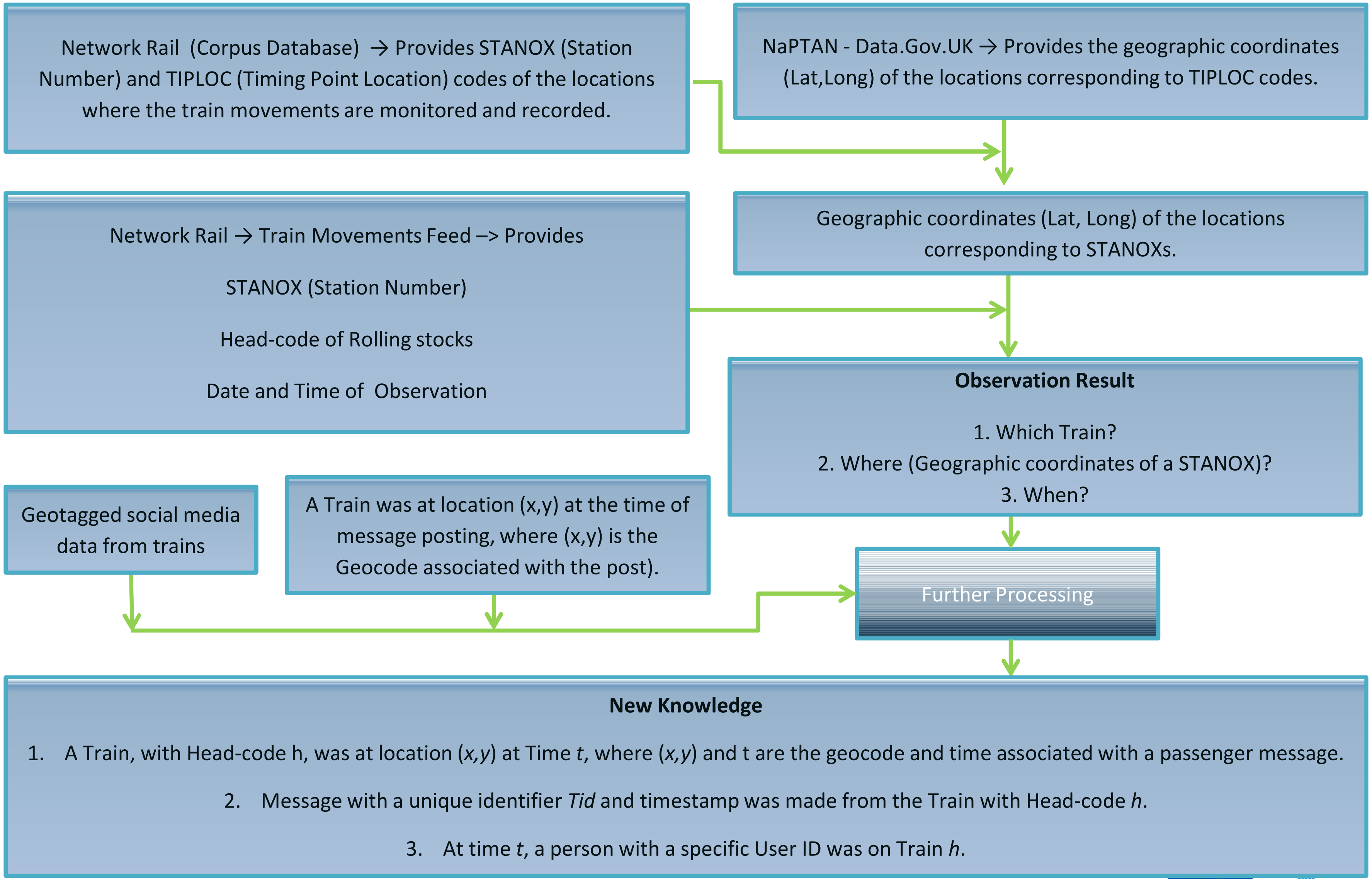
- Leveraging open information on evolving situation from social media
  - Anonymisation and privacy
  - Content mining
  - Geolocation
  - Trustworthiness
- Understanding operational impacts
  - Management of the network
  - Informing travellers



Available Resources

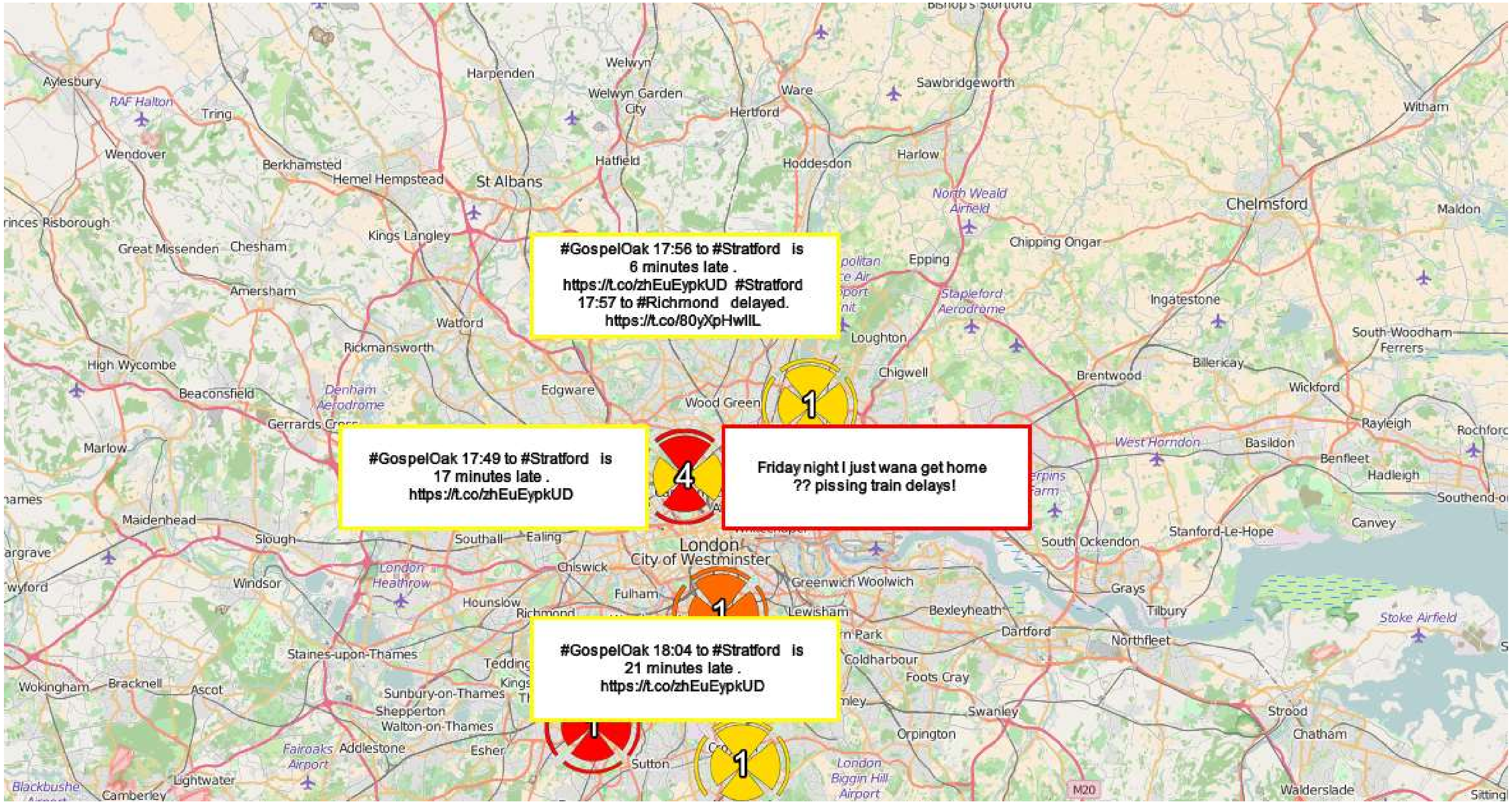
Data Resource	Examples
Social media data	Content, geolocation, time of creation, links to other content
IM public data	Live vehicle movements, train describers, notifications of TSRs etc.
Ordnance survey	Infrastructure layout
ATOC data	Timetables, fares and supporting information
NaPTAN	Information on access points / interchanges

Data Fusion Approach



Geolocated, Grouped Incidents

Capacity for Rail



Deliverable D3.4.2

Capacity for Rail

Summarised in D3.4.2 - “Verified Data Architecture”

- Key topics:
  - Development of existing models (multimodal concepts and the ON-TIME data dictionary)
  - Models and architectures for handling diverse dynamic data resources (RaCoOn and associated implementation best practice)
  - Linked Open Data (open data in support of multimodal operations, link to SP4 and sensor data)
  - Crowdsourced data for improved situational awareness

SEVENTH FRAMEWORK PROGRAMME

Thank you for your kind attention

John Easton

University of Birmingham

j.m.easton@bham.ac.uk