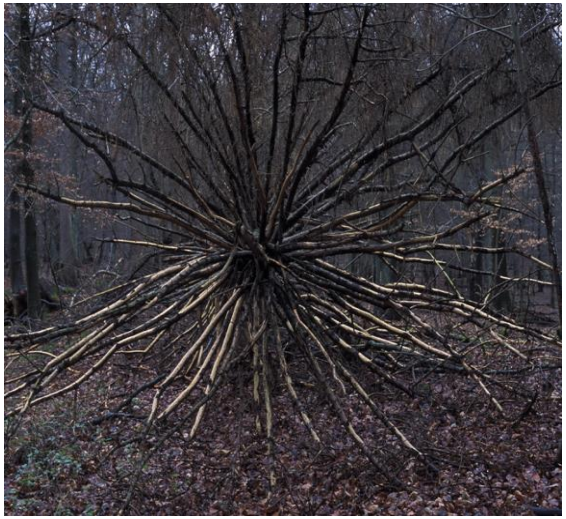
Data to Intelligence Research Program

DIGILE



is about
focusing on **Big Data**, data reserves
and user-centric service development
www.datatointelligence.fi

Photo: Big Bang, December 2006 by Vesa Vähälummukka

Data to Intelligence Program Implementation Facts

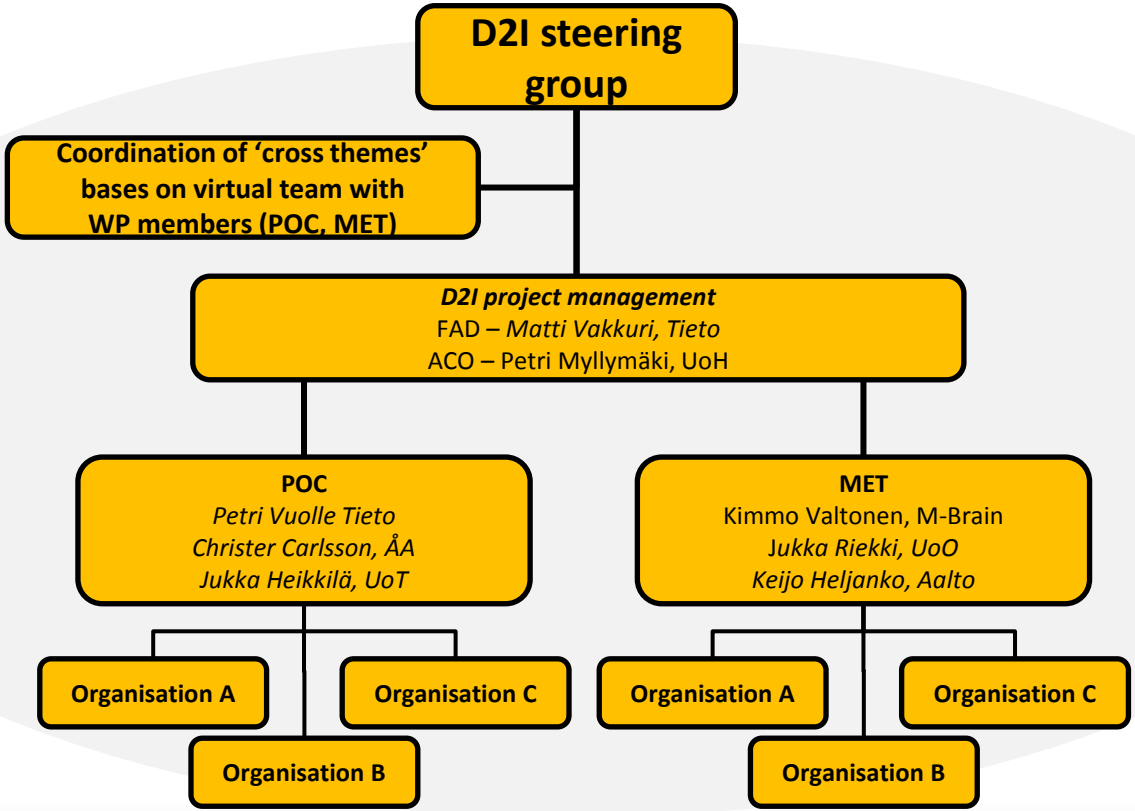
DIGILE

- Driving company: Tieto
- Whole program (estimate)
  - 2012-2015
  - Annually (plan when program started)
    - 15 large companies
    - 10 Small & midsize companies
    - 10 research institutes
- 2014
  - 44 companies
    - 25 large
    - 19 Small & midsize
  - 14 research institutions
  - Work Packages:
    - POC – Proof of Concepts
    - MET – Enabling Methodologies

340 person years
39,5 M€ total cost
60/40 funding model
16,0 M€ company funding
19,5 M€ TEKES funding

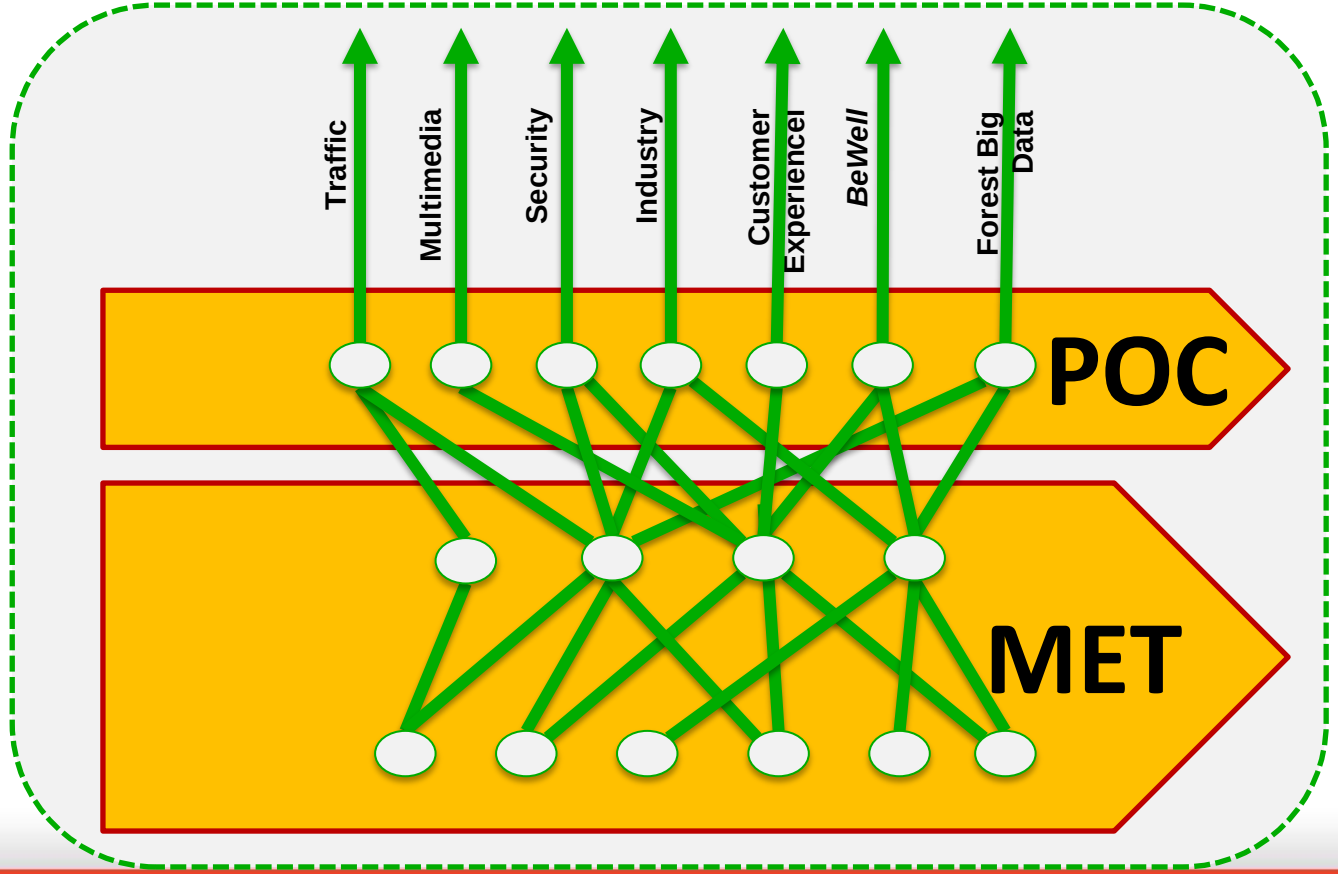
65 person years
11,8 M€ total cost
60/40 funding model
7,02 M€ company funding
4,68 M€ TEKES funding

D2I Program Organisation 2014



The D2I development bases on common D2I framework with interoperable research themes

DIGILE



POC = Proof of Concepts

MET' =Enabling Methodologies

DATA TO INTELLIGENCE - MAIN RESULTS OBTAINED

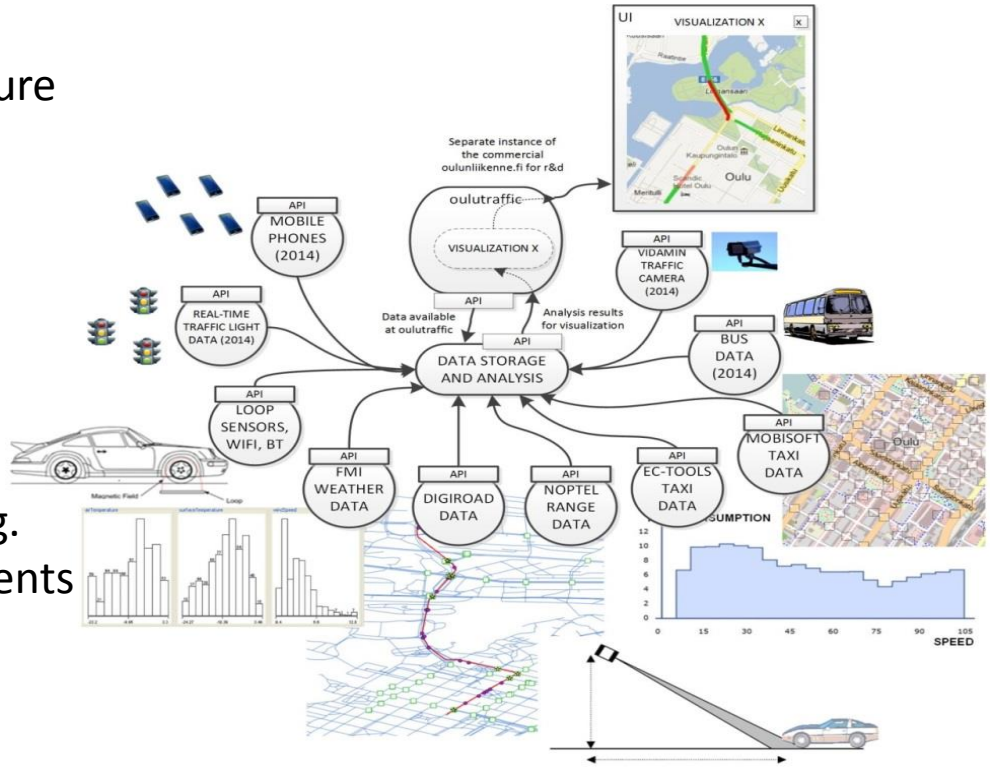
DIGILE

- **Multimedia**
  - The first prototype Predictive browser - Provides faster and more accurate recommendations
  - Sensory pipelines for context-aware computing
  - Data intelligence meets consumer - Smart Image Service prototype client under Trial testing
- **Security**
  - Cleansing of personally identifiable information
  - Web Content Classification
  - Security Data model
  - Visualization of accidents - Gravity
  - Investigation and testing of PRONTO database
- **Traffic**
  - Rich traffic-related data set
  - Traffic simulation tool
  - Braking distance application
- **Industry**
  - Intelligent on-site services
  - Intelligent Worksite Solutions
  - Remote services
- **Customer intelligence**
  - Context-Aware Predictive Modeling of Bus Ridership
  - Context intelligence for marketing
- **BeWell**
  - Human well-being
  - Young elderly
- **Forest Big Data**
  - National forest asset management

Oulu Traffic Pilot

DIGILE

- The Traffic Pilot offers a situation picture from Oulu Region
- The beta version was published in September 2013 at http://beta.oulutraffic.fi
- Information is offered via Web clients and users can select the type of information they are interested in, e.g. fuel consumption in interesting segments

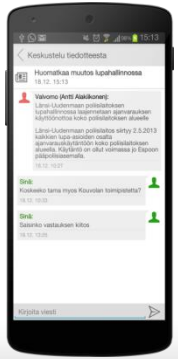


The prototype of low barrier solution for near miss security and safety incident collection

DIGILE

Prototype solution to collect security and safety related incidents with easy to use mobile solution is created. Main functionalities of the mobile prototype:

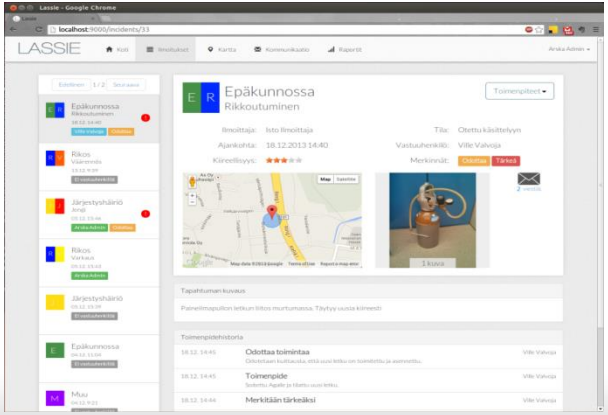
- Easy way to report the incident
  - Classification, description, urgency
  - Automatic time, date and location
  - Photo attachments
  - Edit functionality for adding more information later
- Communication with security and safety organization
  - To get additional information about incident if required
  - To inform/warn mobile users about possible threats



Solution also includes web based user interface for security and safety organization to handle the incidents. Main functionalities of the web user interface:

- Incident processing
  - Processing responsibilities
  - Actions and follow-up
- Communication with mobile and web users

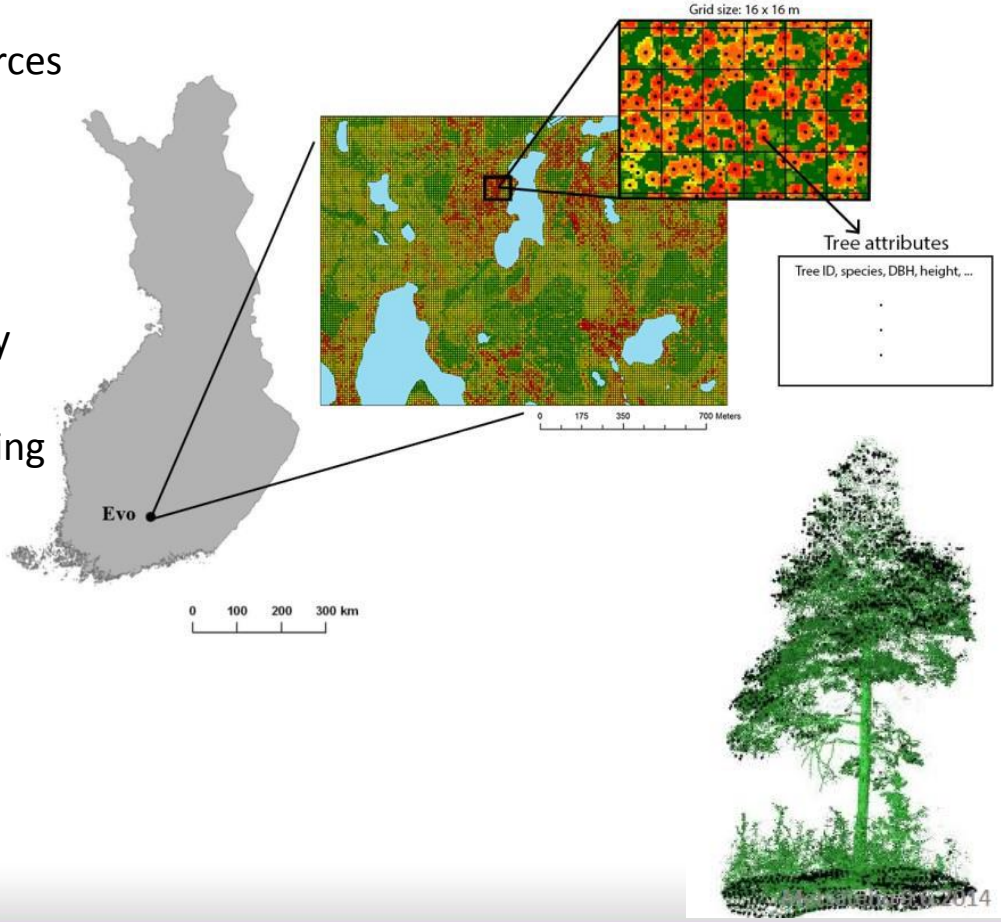
Prototype will be piloted during 2014 and it will be used as a platform to present and visualize security and safety situation information created at D2I Security Ecosystem.



Forest Big Data

DIGILE

- Comparison of all relevant RS data sources in the same study area
- Tree mapping accuracies with 2D/3D TLS/MLS in various forest
- conditions
- Accuracies of the structural tree quality attributes such as tree species,
- stem curve and timber assortments using TLS and MLS
- Preharvest measurement accuracies in mature well-managed stands
- using high-density ALS and DSI data
- Methods to link TLS, MLS and ALS data together



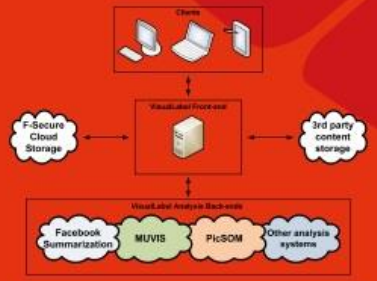
DATA TO INTELLIGENCE (D2I)

VisualLabel:
OPEN ANALYSIS PLATFORM FOR IMAGE AND VIDEO

KALLIO JARNO, RANTANEN PETRI, SILLBERG PEKKA
(kallio@pv.com) (petri.rantanen@tut.fi) (pekka.sillberg@tut.fi)

Goals:

- Design integration specification for content analysis back-ends (MUVIS, PicSOM) and cloud storage platforms
- Provide a platform for feasibility and UX testing
- Primary focus on photos, with the possibility to expand to audio and video content
- Implement a media analysis service as a part of VisualLabel project of the D2I program
- Deploy as a FORGE service



System Features:

- Front-end services implemented using free and open source components
- Search by both content similarity and tags (keywords and people)
- Automatic tag generation using analysis back-ends
- Support for connecting 3rd party content storages (e.g. Picasa, Facebook)
- Direct and indirect user feedback for generated tags and search results

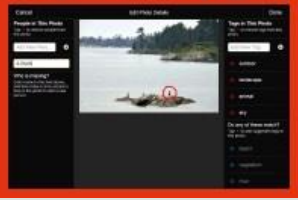
Web User Interface:

- Used in concept evaluation
- Demonstration of core features

1. Search by people and keyword tags

2. Similarity search with direct user feedback

3. Edit tags, view automatic tag suggestions



Scalability of a Cloud-Based Data Store: Improving HBase performance

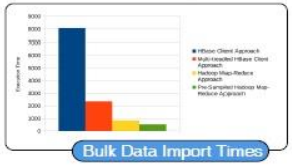
Aalto University
Mario Cerdan, Keijo Heljanko, Rıdvan Döngelci
firstname.lastname@aalto.fi
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

- We analysed and experimented with HBase NoSQL database to observe and improve its performance on video metadata.
- HBase is optimized for write heavy workloads. HBase is highly configurable and should be configured aggressively to get optimum performance.
- Java virtual machine, garbage collector, Memstore, compression and Hadoop parameters are crucial for HBase performance.
- Bulk data import to HBase is fairly important task for various purposes and should be done efficiently.

Summary

- Following four approaches are investigated for bulk data import to HBase:
  - **HBase Client Approach**
  - **Multi-threaded HBase Client Approach**
  - **Hadoop Map-Reduce Approach**
  - **Pre-Sampled Hadoop Map-Reduce Approach**

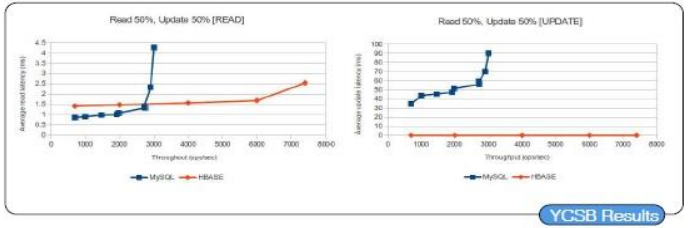
Bulk Data Import Approaches



Bulk Data Import Times

- We further compared HBase performance with MySQL performance to analyse benefits of NoSQL database.
- The benchmarking is based on industry standard Yahoo! Cloud Serving Benchmark (YCSB).
- 50% read and 50% update workload demonstrates that HBase is superior to MySQL with respect to updates and fairly competitive with respect to read performance.

HBase Performance Benchmarking

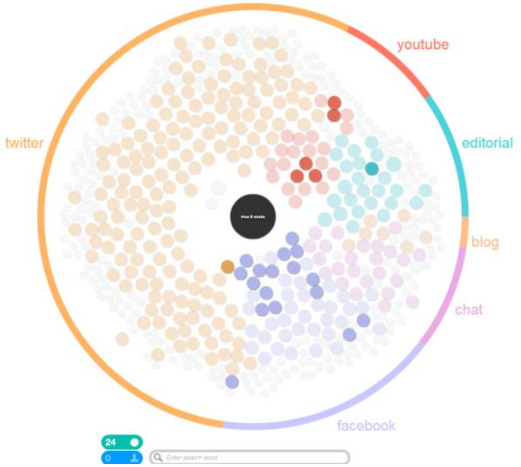


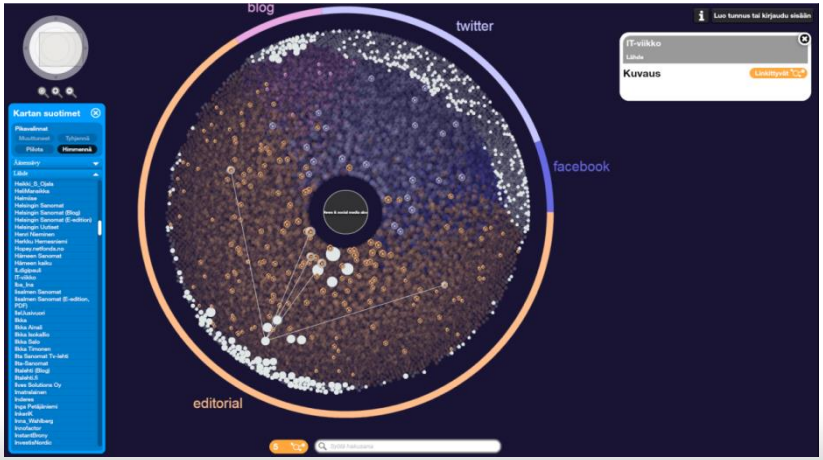
YCSB Results

Context intelligence for marketing

DIGILE

- Text analysis combined with visualization of the entire business-relevant media landscape:
  - study of sources
  - industry topics, companies and
  - brands in a single view

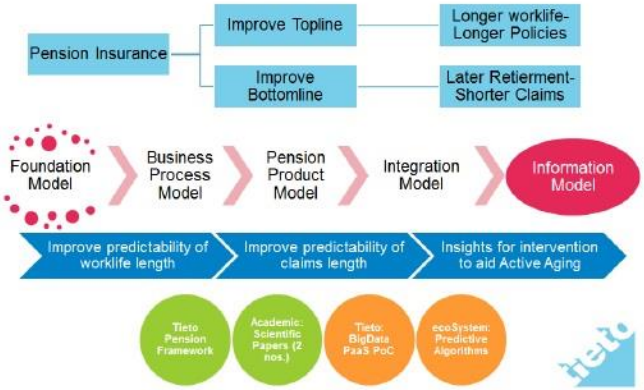




DIGILE | DATA TO INTELLIGENCE

BeWell Ecosystem

Pan-Nordic Pension Framework Pilot



Status

- Pension Foundation Model is complete for Finland — DONE
- Pension Business Process Model is partly completed for Finland — OnGOING
- Research for mapping the Pension Insurance landscape is in final stages — OnGOING
- Research for identifying Forms of Life around Retirement is underway — OnGOING

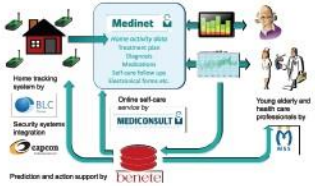
Tieto, ABO, UTU, VTT, JYU

DIGILE | DATA TO INTELLIGENCE

Home Activity of Young Elderly

BeWell Pilot (started Feb 2014)

Maintaining quality of life, ensuring safety at home, promoting independence in daily activities



Key Objectives Pilot Phase (2014)

- Create the Proof-of-Concept:
  - Conduct the pilot phase with 10 young elderly participants (Life based design)
    - installing home tracing system, which transfers sensor data to safe servers
    - monitoring activity patterns for 12 weeks (starting Sept) and providing activity data for partners for further data fusion
    - engaging participants in co-design sessions and involving them for further planning (interviews, workshops)
  - Preliminary analysis of data, methods evaluation and development
- Use ICF (International classification of functioning, disability and health) to structure the collected everyday life activity data
- Piloting the usage of Business scrum-method in creating a new way to co-operate within a heterogeneous group

Key Objectives R&D Phase (2015->)

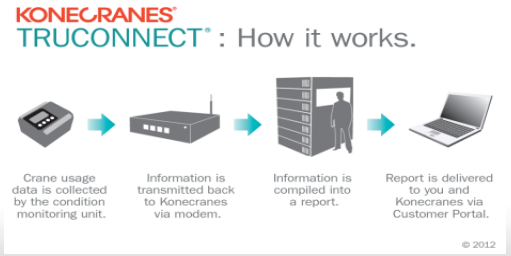
- Provide models for predicting daily behaviour and activity patterns
- Conduct further analysis of activity patterns to detect e.g. early memory disorders
- Provide data via home tracing system on changes in daily activity patterns
- Develop a virtual homecare forum, which enables data sharing and interaction between young elderly, relatives and professionals
- Carry out co-design activities with young elderly and other stakeholders

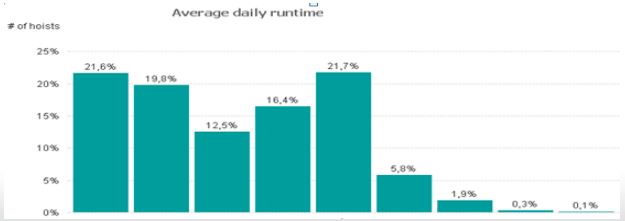
VTT benefe BLC capcon MEDICONSULT

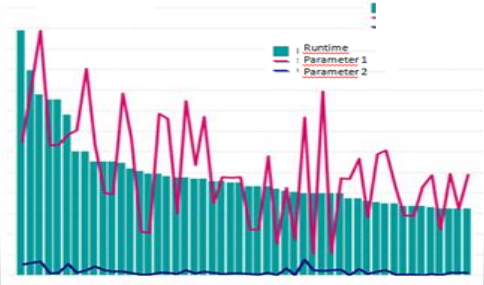
Remote and highly automated services

DIGILE

- Business concept finalized – support Industrial Internet
- Lifecycle monitoring of key components and detection of crane use with sensors
  - Development of crane usage, performance and condition monitoring methodology
  - Supports equipment lifetime management and maintenance optimization
- Sensor networks and single sensors connected to the Remote services to provide field data.
- The result is a concept supporting equipment fleet analysis and lifetime management, maintenance optimization, field service operations, and customer reporting.







The Importance of Big Data for the Finnish and Global Economy

DIGILE

- **Revolutionizing Healthcare**
  - Better healthcare, better medicin, telemedicin, better quality medical diagnosis, less error, longer life, direct utilization of human DNA
- **Better results from Education & Learning**
  - Better tools for institutions to improve the outcomes for each individual student
- **Smarter and more effectiveTransportation & Logistics**
  - Smarter, more automated transportation, cars, trucks, rail traffic, sea transportation, minimized energy consumption, optimized routes, less pollution, less traffic jams
- **New Paradigm for Energy consumption, production and delivery**
  - Real-time monitoring and control, analysis, consumption forecasting on individual level as well as for societies and nations. Less energy spill, less waste and pollution. Better utilization of natural resources.

The Importance of Big Data for the Finnish and Global Economy

DIGILE

- **Improved and more sustainable everyday producs around us**
  - Products become more intelligent and aware of the users and circumstances where they are being used
  - New products make life easier
- **Safety & Security**
  - Less crimes, less terrorism, less accidents, better defence
- **Banking and Financial Services**
  - Real-time economy
  - Better management of financial risk
  - Better financial analysis and forecasting, anti money laundering
  - Better financial products through understanding the customer needs.

DIGILE

DATA TO INTELLIGENCE

Innovative data intensive services and business development

www.datatointelligence.fi