Reusable, Sharable, and Transferable Smart Data Platform for Collaborative Development of Data-Driven Smart Cities

Background :

- UN's Sustainable Development Goals (https://sdgs.un.org/goals)
  - Goal 11: Make cities and human settlements inclusive, safe, resilient and sustainable
  - Goal 13: Take urgent action to combat climate change and its impacts
  - Goal 17: Strengthen the means of implementation and revitalize the global partnership for sustainable development
- Society 5.0 (https://www8.cao.go.jp/cstp/english/society5_0/index.html)
  - "A human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space."

Targets:

- Disaster risk reduction and sustainable transport
  - Smart environmental Tourism and Sustainable Mobility (Dalat city, Vietnam)
  - Smart Dengue Early Warning System (Cauayan City, Philippines)
  - Smart Outdoor Activities (Singapore)
- Atmosphere and Climate Change
  - Transboundary Air-Pollution Forecasting (Brunei)
- Multi-stakeholder partnerships and voluntary commitments
  - xData Platform and Event Data Sharing: A decentralized and collaborative approach to fast, economically, and sustainably develop user-centered applications.

Speaker: Minh-Son Dao

Reusable, Sharable, and Transferable Smart Data Platform for Collaborative Development of Data-Driven Smart Cities

Project Members :

- Minh-Son Dao, *Senior Researcher, NICT, Japan (PI)*
- Asem Kasem, *Assistant Professor, Universiti Teknologi Brunei (UTB)*
- Dang Thanh Hai, *Lecturer, Dalat University, Vietnam (DLU)*
- Filip Biljecki, *Assistant Professor, National University of Singapore (NUS)*
- Betchie Aguinaldo, *Associate Professor, Isabela State University, Philipplines (ISU)*

Project Duration :

- 24 months (2019/04/01 – 2022/03/31)

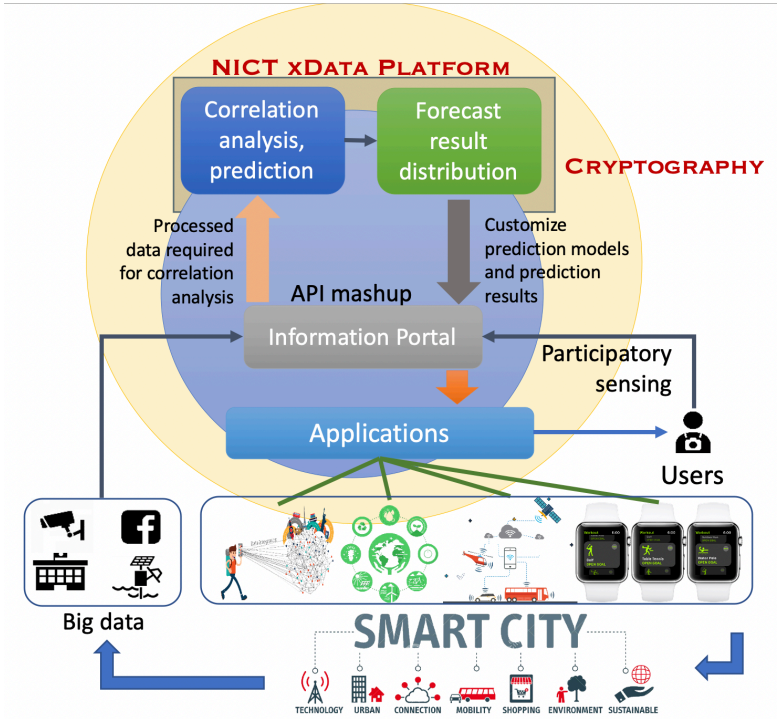
Project Budget:

- 40,000 USD/year

Project Activities: Scientific and Technological Development

1. Scientific

- Values and Complex Events Prediction
- Insights from cross-data
- Decentralized and collaborative development
- Data Visualization and Geo-based Navigation



2. Technological development

- CRNN-based models for (Transboundary) Air Pollution Prediction using peripheral and weather from stations and open data.
- 3D-CNN models for complex event prediction and detection using spatiotemporal cross-data.
- Partial Periodic Spatial Patterns Discovery in Spatiotemporal Databases
- ETL APIs (e.g., data loader) for supporting data collection and sharing/exchanging among devices (e.g., mobile sensor boxes, lifelog cameras, peripheral stations), xData Edge (i.e., local server) and xData Platform (i.e., basement infrastructure)
- Federated Learning and Transferred learning versions of CRNN-based models for decentralized and collaborative development different instances of original models.
- Data collection system (e.g., sensor networks, crowdsourcing)
- A risk-avoidable navigation system based on cross-data

Project Activities: Experiments including field testing

- Sensor Networks designation and deployment
  - Environment-mobility sensor networks in Dalat city (Vietnam)
  - Dengue sensor networks in Cauayan city (Philippines)
- Data structure and database designation based on 3D-GIS data format to harmonize data exchanged among partners and to further adapt to smart cities.
- Data collection:
  - Collect haze trans-border-related datasets (air pollution, fire forest, weather) from ASEAN countries
  - Collect Dengue-related dataset in Cauayan city (Philippines)
  - Organize Tokyo Olympic (Japan) and Dalat city (Vietnam) data collection campaigns to collect peripheral sensing data, lifelog data (images and questionnaire), and open data. The campaign has happened for three months (September-November 2020)
- Transfer knowledge:
  - Transfer prediction models and association rules discovery to partners for reusing on local data from NICT to UTB and DLU
- Models Evaluation:
  - Evaluate the adaptability and accuracy of transferred models on local data (UTB, NICT, DLU, NUS)
- System Manual Instruction:
  - Construct the website to give manual instructions for reusing and sharing products developed by partners (https://www.xdata.nict.jp/xDataPFDocs/WebAPI/1.0/en/)

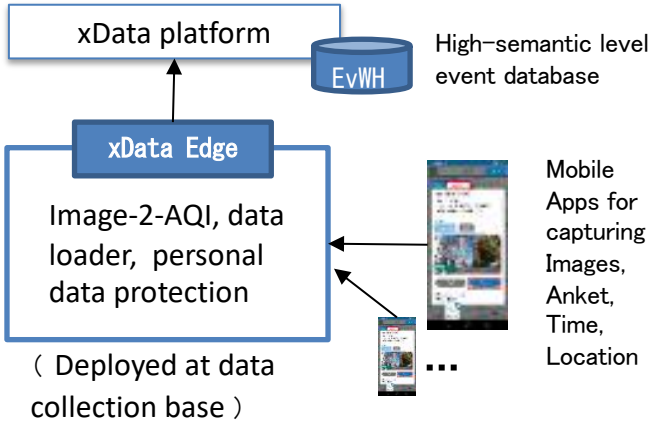
Project Activities: Budget plan in detail

1. 20,000 USD for renting GPU cloud (for high-performance and collaborative training deep-learning models) and GIS cloud (for flexible and instant crowdsourcing). These resources are shared among partners.
2. 20,000 USD divided equally for four participants for supporting local sensor networks setup (Dalat city - Vietnam, Cauayan city – Philippines, Brunei), data acquisition (all partners), and miscellaneous things facility (all partners) towards enhancing the project activities.
3. All traveling and on-site events organizing, and attending are canceled due to the Corona pandemic. Hence, no budget is located for this category.

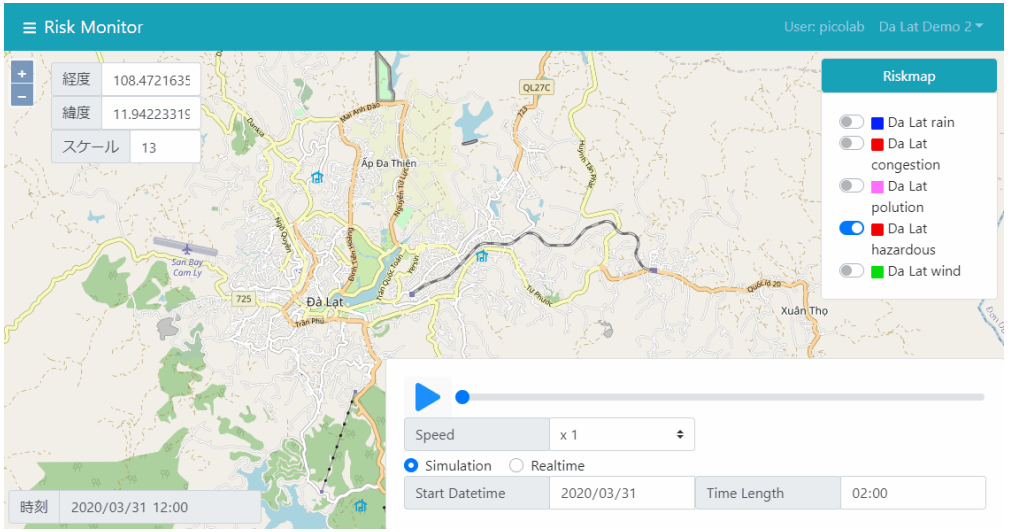
R&D results: Scientific and technological

1. Values and Complex Events Prediction
   - MASTGN: Predict air pollution using Multi-Attention Spatio-Temporal Graph Networks on Tokyo peripheral dataset (air pollution, weather)
   - Image-2-AQI: Predict environmental quality using image retrieval and deep learning models on lifelog images, times, and locations.
   - 3DCNN: Predict events using CNN, raster images, and concurrent fusion on cross-data (weather, congestion, accident, and tweets).
2. Insights from cross-data
   - Mining of Spatial High Utility Itemsets in Very Large Spatiotemporal Databases, applied for Tokyo and Kobe dataset. The patterns can help to predict the length of congestion in future.
3. Decentralized and collaborative development
   - Transferred Learning: The CRNN AQI prediction model is re-trained on Brunei dataset to predict PM10.
   - MM sensing system: the integration system to connect personal multimedia devices, xData Edge (local servers) and xData PF for predicting events (e.g., AQI, congestion)
   - Data loader APIs: many data sources are uploaded and shared by different data loader functions designed according to related data types.
4. Data Visualization and Geo-based Navigation
   - A risk-avoidance navigation system based on localized torrential rain data to plan the optimal route for travelers in real-time mode.

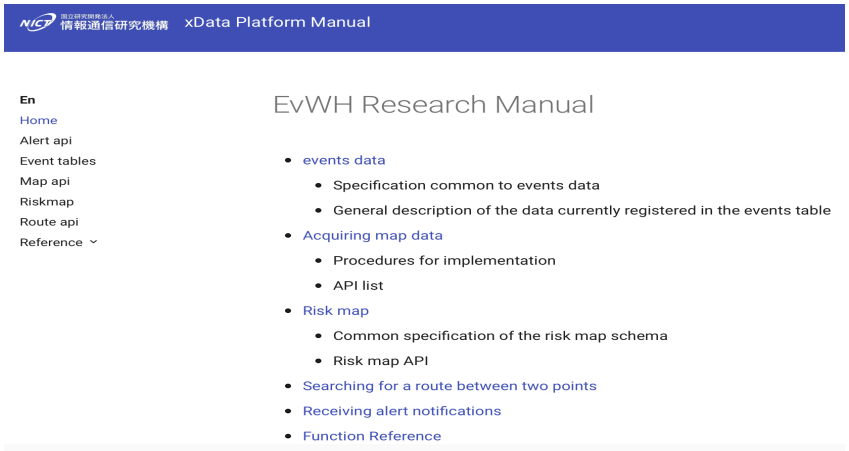
R&D results: Application (or system) development



MM sensing system

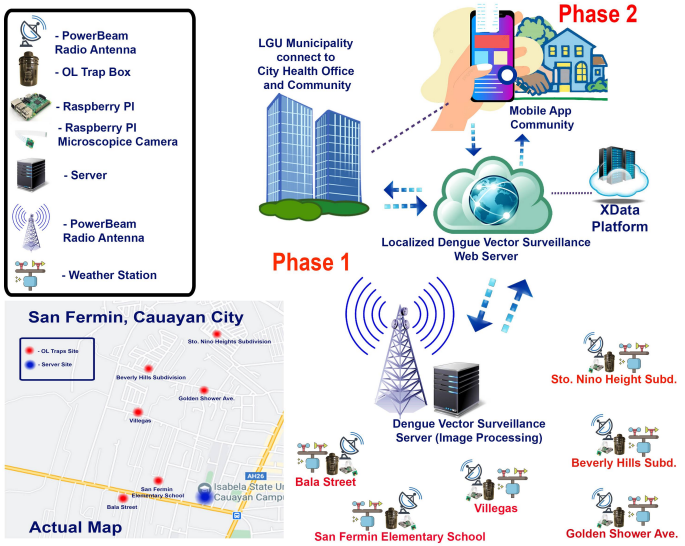


Dalat navigation risk map



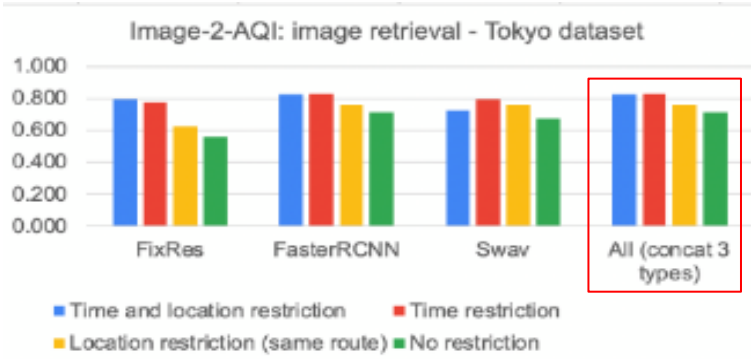
APIs library and manual instruction website (included all developed predcition models, association discovery, data loader, and risk map)

https://www.xdata.nict.jp/xDataPFDocs/WebAPI/1.0/en/

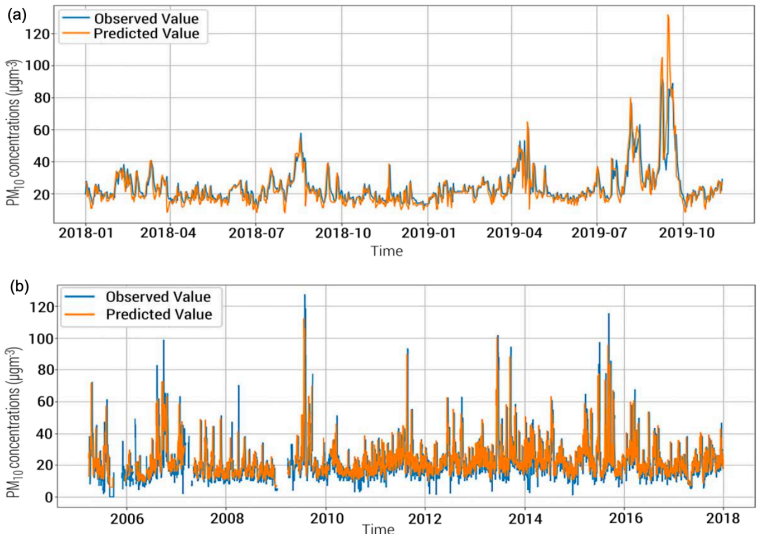


Dengue Early Warning System Architecture

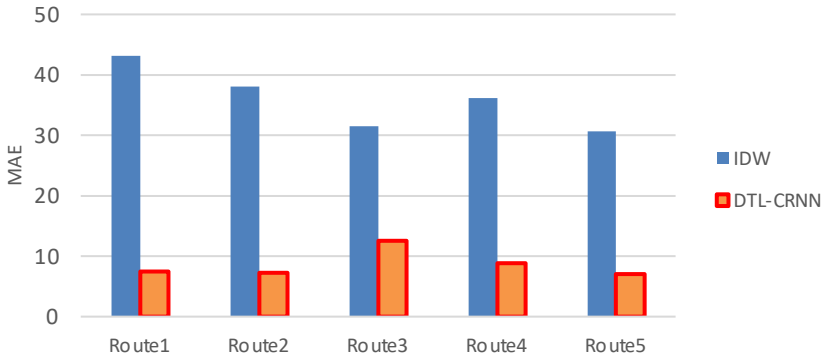
R&D results: Experiments including field testing



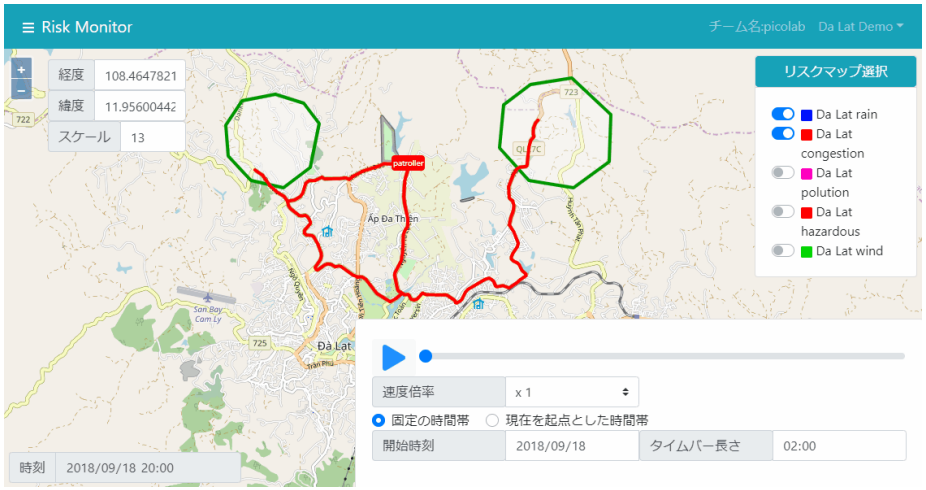
Prediction accuracy in evaluation experiments using data collected by Tokyo Dataset.



Transferred CRNN model predicts values of PM10 concentrations at Mumong station, Belait district, Brunei, for (a) testing sets (2018-2019), and (b) training sets from 2005 to 2017



CRNN model: 15-37% improvement in prediction error compared to spatial completion method (IDW method) of measurement station data (Tokyo)



Environment- mobility risk map simulation running on archive dataset of Dala city, Vietnam

Scientific Contribution: Presentations at International Conferences

No	Paper title:	Author names	Affili-ation	Conference name:	The date of the conference	The venue of the conference
1	MNR-HCM Data: A Personal Lifelog and Surrounding Environment Dataset in Ho-Chi-Minh City, Viet Nam	T.L. Nguyen-Tai, D.H. Nguyen, M.T. Nguyen, T.D. Nguyen, **T.H. Dang***	*DLU	Intelligent Cross-Data Analysis and Retrieval Workshop, (ICMR 2020)	26-29/10/2020	Dublin, Ireland (Virtual Conference)
2	ICDAR'20: Intelligent Cross-Data Analysis and Retrieval	**M.S. Dao***, M. Fjeld, **F. Biljecki****, U. Yavanoglu, M.X. Dong	*NICT, **NUS	Intelligent Cross-Data Analysis and Retrieval Workshop, (ICMR 2020)	26-29/10/2020	Dublin, Ireland (Virtual Conference)
3	Leveraging 3D-Raster-Images and DeepCNN with Multi-source Urban Sensing Data for Traffic Congestion Prediction	N.T. Nguyen, **M.S. Dao***, K. Zettsu	*NICT	Database and Expert Systems Applications Int. Conference (DEXA2020)	14-17/09/2020	Bratislava, Slovakia (Virtual conference)
4	MASTGN: Multi-Attention Spatio-Temporal Graph Networks for Air Pollution Prediction	**P.J. Zhapo***, K. Zettsu	*NICT	IEEE Big Data 2020	10-13/12/2020	Virtual Conference
5	Convolution Recurrent Neural Network for Daily Forecast of PM10 Concentrations in Brunei Darussalam	E. N. Aziz, **A. Kasem***, W. S. H. Suhaili and **P. Zhao****	*UTB, **NICT	The 6th International Conference on Low Carbon Asia & Beyond	1-3/9/2020	Shanghai, China (virtual conference)
6	Distributed Mining of Spatial High Utility Itemsets in Very Large Spatiotemporal Databases using Spark In-Memory Computing Architecture	U.K. Rage, S. Ito, **M.S. Dao***, K. Zettsu, C.W. Wu, Y. Watanobe, I. Paik, C.T. Truong	*NICT	IEEE Big Data 2020	10-13/12/2020	Virtual Conference

Scientific Contribution: Social Impacts

1. Open public dataset:
   - SEPHLA: Challenges and Opportunities within Environment-Personal Health Archives
     - 15 publications that use SEPHLA as the benchmark dataset (i.e., citations)
2. Challenges:
   - MediaEval 2020 Insight for Wellbeing: Multimodal personal health lifelog data analysis (https://multimediaeval.github.io/editions/2020/tasks/lifelogging/)
     - 11 teams attended (Vietnam, UAE, India, Pakistan, Rumania, China, Brunei)
3. Conference special session and workshops:
   - ICMR2020 workshop on Intelligent cross-data analytics and retrieval (https://www2.nict.go.jp/bidal/icdar_icmr2020/index.html)
     - 17 submitted papers, 6 accepted papers for oral presentation
   - ICME2021 Special session on Insights for wellbeing: Understand the surrounding environment's impact on human lives (https://2021.ieeeicme.org/conf_sessions)
     - 5 confirm-to-submit papers and more papers are coming (deadline 29/11/2020)

Conclusion:

1. **Scientific and technological**
   - Researched and developed several topics and methods that contribute to the success of the project including air pollution CRNN-based prediction models, 3DCNN event prediction models, periodic frequent patterns mining, predict AQI from lifelog images, safe route navigation planning.
2. **Application (or system) development**
   - MM sensing system
   - A risk-avoidance navigation system
   - Dengue Early Warning System architecture
   - APIs library and manual instruction website
3. **Experiments including field testing**
   - Evaluated (transferred) CRNN-based prediction models in Tokyo and Brunei dataset
   - Evaluated 3DCNN event prediction models in Kobe dataset
   - Evaluated risk map in Dalat city dataset
   - Evaluated Image-2-AQI in Tokyo and Dalat city dataset.
4. **Datasets**
   - Tokyo environment-lifelog dataset
   - Dalat city environment-lifelog-tourism-CCTV dataset
   - Cauayan city Dengue dataset
   - ASEAN air pollution dataset
   - SEPHLA: open dataset for Challenges and Opportunities within Environment-Personal Health Archives

Future works:

1. Scientific and technological
   - Continue ongoing scientific and technological purposes mentioned above
   - Research and Develop more topics and methods focusing on improving the smart of data, the flexible ability of Reusable, Sharable, and Transferable of the system.
2. Application (or system) development
   - Smart environmental Tourism and Sustainable Mobility for Dalat city, Vietnam
   - Smart Dengue Early Warning System for Cauayan City, Philippines
   - Smart Outdoor Activities for Singapore
   - Transboundary Air-Pollution Forecasting for Brunei
   - Open dataset for open science data community
   - Completed xDataPF – xDataEdge and MM sensing system with high privacy protection
3. Experiments including field testing
   a) Field experiment of environmental quality data collection by residents using MM sensing and customization of short-term prediction of environmental quality of tourist spots (Dalat City, Vietnam) and environmental health (Cauayan City, Philippines) using collected data.
   b) Field experiment of environmental walking (Tamagawa Walking Society, Tokyo) evaluated the effectiveness of environmental quality data collection in various environments such as highways, rivers, and residential areas.
   c) Hackathon to attract the attention from industry-academy-government on environment-human topics, expected to be organized in NUS