



# How to organize for AI? An interview with Yao-Hua Tan

Rainer Alt<sup>1</sup>

Published online: 23 October 2021

© The Author(s) 2021

## Personal details

Yao-Hua Tan is a professor of Information and Communication Technology at the ICT Group of the Department of Technology, Policy and Management at Delft University of Technology and program director of the Customs and Supply Chain Compliance master of the Rotterdam School of Management of the Erasmus University Rotterdam. He was also Reynolds visiting professor at the Wharton Business School of the University of Pennsylvania. His research interests are service engineering and governance, ICT-enabled electronic negotiation and contracting, multiagent modeling to develop automation of business procedures in international trade. He is scientific coordinator of the EU-funded research project PROFILE on information technology (IT) innovation such as blockchain and data analytics and artificial intelligence (AI) methods for advanced risk assessment to control international supply chains. Yao-Hua has been an editorial board member of *Electronic Markets* since 2006 and joined *Electronic Markets*' advisory board in 2020. In the interview with Rainer Alt (left picture), Yao-Hua (right picture) reports on his experiences in AI projects.



✉ Rainer Alt  
rainer.alt@uni-leipzig.de

<sup>1</sup> Information Systems Institute, Leipzig University, Grimmaische Str. 12, 04109 Leipzig, Germany

## Background details

International trade is a highly interorganizational domain that impressively shows the intricacies of networked business. The European Union alone comprises more than 2 million exporting and importing organizations in the 27 member states as well as over 400 customs offices (Henningsson et al., 2011). Although IT has been applied with tracking, documentation and EDI systems since the late 1960s, the fragmented nature of this industry has also led to heterogeneous information systems that embody heterogeneous process and data definitions. For example, an analysis conducted by the container shipping company Maersk in 2014 showed that a refrigerated container filled with roses and avocados from Kenya to the Netherlands involved 100 people in over 30 organizations and led to over 200 interactions. The entire journey required some 34 days, whereof ten days passed with the container sitting and awaiting for documents to be processed, which were repeatedly printed and re-entered manually (Park, 2018). In addition to inefficiencies like waiting time, these manual tasks were error-prone and the case revealed that one critical document even went astray. Similarly, another example showed that “a truck with export goods crossing the Russian-Finnish border may be required to present as much as 40 different paper documents to be granted passage” (Henningsson et al., 2011, p. 2). Among these documents were packing lists, bills of lading, letters of credit as well as various customs documents, such as import declarations.

Customs is an important actor in these cross-border transactions. A main responsibility is to assess the compliance with trade and safety regulations, which involves determining customs duties, preventing fraud and detecting threats such as associated with narcotics and explosives. However, the large portion of manual processes and a growing trade volume have made the goal to inspect all incoming and outgoing cargoes an illusion (Heijmann et al., 2020). For the Netherlands alone, the customs authority processes an estimated amount of comprised some 160 million declarations annually, a figure that and is expected to increase to over 500 million by 2022 (Segers et al., 2019). Multiple approaches have taken place to address this apparent dilemma in many countries and

important projects funded by the European Commission were conducted in joint collaboration between authorities, businesses and research institutions (see Table 1). They emphasize the role of IT in establishing an interoperable and secure information infrastructure among the participating actors as well as in devising functionalities that automate customs processes and pave the way towards approaching the 100 percent inspection goal.

From a technological perspective, future solutions feature the growing convergence of several general purpose technologies (Alt, 2021). Among the examples are blockchain technologies that create new distributed infrastructures with automated procedures based on smart contracts (Segers et al., 2019), the use of internet-of-things (IoT) and mobile technologies such as smart containers and drones (Heijmann et al., 2020) as well as the application of various forms of data analytics to automatically detect anomalies and the like (Rukanova et al., 2021). At the same time, the projects have shown that IT is a necessary enabler, but not a sufficient ingredient for successful interorganizational process (re)design. In their attempt to devise solutions in real-life settings, the projects shown in Table 1 have pursued the living lab setup (Stijn et al., 2009) and illustrated that technological change needs to coincide with organizational and institutional redesign (Tan et al., 2011). For example, trusted information infrastructures allow process changes like the trusted trader concept that implies a change from push to pull principles. Similarly, it is important to understand work practices (e.g. in risk assessment) when developing data analytics models. Observations in these respects are the topic of the following interview.

## What is your observation on research in the domain of data analytics and AI?

My impression is that many researchers are active in the field of data analytics and AI and that research in this area already has a long tradition. Today, we see that substantial knowledge on large databases, powerful algorithms and data processing are available. Indeed, plenty of research exists that reports on these developments and shows how they perform. In most cases, the results are positive and convincing, for example, on the superior results of applying algorithms in the areas of fraud detection or other anomalies as well as on how to choose the “right” algorithm.

## How do you see the situation in practice?

In my research, we are extensively collaborating with customs and other border inspection agencies such as product- and food safety inspections and international businesses from many industry sectors. This research dates back to the time of electronic data interchange (EDI) and today comprises many projects, also large projects funded by the European Commission (EC), such as ITAIDE, CASSANDRA, CORE and now PROFILE (see Table 1) where the focus is on developing data analytics to improve customs fraud and security threat detection. My impression from several projects is that despite organizations have “beefed up” their technological infrastructure for AI, the quality of data in these data lakes is often very low. The pilot projects confirmed that the algorithms worked as prototypes in a lab environment, but once

**Table 1** Overview of projects in the international trade context (Source: cordis.europa.eu)

| Project               | Description   |
|-----------------------|---|
| ITAIDE (2006–2010)    | <ul style="list-style-type: none"> <li>● Acronym: Information technology for adoption and intelligent design for e-government</li> <li>● Goal: Development of innovative solutions for e-government in the context of cross-border trade</li> <li>● Results: Design of a common information model for electronic documents and document mapping software</li> </ul>   |
| CASSANDRA (2011–2014) | <ul style="list-style-type: none"> <li>● Acronym: Common assessment and analysis of risk in global supply chains</li> <li>● Goal: Supply chain visibility and reuse of trade information along the entire chain to enable business operations and cross-border security inspections</li> <li>● Results: Design of a data pipelines as basis of dashboards to enable secure data sharing in heterogeneous system environments</li> </ul>                               |
| CORE (2014–2018)      | <ul style="list-style-type: none"> <li>● Acronym: Consistently optimized resilient secure global supply chains</li> <li>● Goal: Integration of real-time optimization tools as well as security and tracking data into supply chain operations to ensure process efficiency and reduced environmental impact</li> <li>● Results: Design of a security and supply chain resilience ecosystem for international trade based on trusted worldwide networks</li> </ul>    |
| PROFILE (2018–2021)   | <ul style="list-style-type: none"> <li>● Acronym: Data analytics, data sources, and architecture for upgraded European customs risk management</li> <li>● Goal: Uptake of state-of-the-art data analytics and incorporation of new data sources for more effective and efficient European customs risk management</li> <li>● Results: Design of a EU-wide risk data sharing architecture and of tailored solutions for customs risk management based on AI</li> </ul> |

we tried to scale up these prototypes and integrated them in large-scale operational systems in customs organizations and international companies, the results were disappointing. Our results indicated that the problem is largely a business-IT alignment issue. Many organizations tend to outsource AI development, and data scientists then apply their algorithms to the data provided by the government agency or commercial companies. However, these AI specialists then lack the specific domain knowledge on what specific parameters to focus on, and are unable to judge which parameters are actually relevant to increase the performance of a specific business process. Let me illustrate this with one example: An organization employed a team of 150 data analysts who were very advanced in the field of data analytics. They produced impressive prototypes, and statistically significant relations, but it turned out that of the vast majority of these relations were completely useless when assessed by the domain experts—in fact, every domain expert could quickly determine that the analyses were useless spurious relations and often the result of overfitting of biases in the data set. Furthermore, it turned out that it took these domain experts more effort to weed out the spurious relations, than finding the relevant parameters and relations completely manually from scratch. In the end, the data analytics unit was dissolved and integrated more in the existing organization.

### **You recommend that domain experts need to be “in the boat” early on?**

This is exactly the point. We observed that in AI and business analytics one should spend at least as much effort on the data preparation, such as selecting a data set that has the least biases as well as cleaning (i.e. removing extreme outliers and incorrect data) and identifying the relevant parameters in the data set, as effort spent on developing the actual algorithms. This data preparation cannot be done by technical specialists, because they lack the necessary domain knowledge and business background. Once the training data set is well-prepared, applying the data analytics and AI technology to these data sets afterwards is relatively straightforward and easy. The real challenge is the data preparation phase and domain expertise is vital for this task. Therefore, one of our key learnings and recommendations is that data analytics or AI teams should at least consist half of domain experts, who are able to “feed” the technological specialists with the proper data sets. It is not sufficient to just throw big data algorithms to arbitrary data sets. More important is the question, how to make sense of large data sets. I have often encountered the saying that “the more data you have, the better results you get with machine learning”, but I am convinced that this is a myth. If quality of the data is garbage, then the analytics

results will be garbage too. It is important to know how the “real world” works and you will see that only a very carefully composed liaison of technological and domain experts prevents the disappointments and leads to insights that are valuable for the business.

### **How do you see the topic in the inter-organizational environment?**

First, we see that large organizations are developing their own data lakes. This is already challenging since in global organizations different business units in different countries need to participate. Almost like in the inter-organizational system setting, they have non-standardized heterogeneous data and processes, which requires much mapping effort to harmonize and standardize data into one suitable data lake. This is a very similar issue to what we know from EDI projects since the 1980s. EDI has been applied in international trade from the beginning, and to create robust analytics for such transactions, semantic interoperability by standardizing message data is a valuable approach.

### **Do you see other promising approaches in this direction?**

Absolutely, there are initiatives in various countries that aim at improving the semantic interoperability of data in cross-country supply chains and at addressing the problem of data preparation in multi-organization supply chains. Among the examples are the Industrial Data Space project (IDS) by Germany’s Fraunhofer institute and the creation of a European federated data infrastructure called GAIA-X as well as the standards that are currently developed by the Digital Transport and Logistics Forum (DTLF) and activities from the EC in the directorate-general for mobility and transport as well. We are participating ourselves in a large EC-funded project called PROFILE together with companies such as IBM, the Dutch Institute of Applied Science (TNO) and many national customs administrations (e.g. in the Netherlands, Belgium, Sweden and Norway) to develop advanced data analytics for compliance risk management of international trade. We observed that organizational issues to apply AI in the customs context amounted to 80% of all effort and that only 20% of this effort was technical in nature.

### **What implications do you see for the customs sector?**

Traditionally, the customs sector is faced with handling documents from various parties. These documents were supported by EDI early on and standards such as UN/CEFACT

emerged. However, in parallel the need not only to transfer this growing volume of transaction data among many participants is challenging, but also to meet the regulatory obligations of containing fraud and other misbehaviors. Let me illustrate this with one figure: In 2020, the volume of customs declarations in the Netherlands amounted to 180 million. For 2021, the forecast is a volume of 800 million, and expected to grow even more in the following years. This increase is mainly due to the steep increase in e-commerce sales. Although customs agencies have a long tradition in collaboration among each other, there is an acute need to automate inspection activities with data analytics and AI, and to develop that jointly with groups of customs administrations. This AI innovation collaboration happens not only between EU member states, but also across regions, for example between EU member states and China, US, Middle-East, Latin America or Africa.

### Can you generalize your observations to other application areas?

Like with EDI, the application of data analytics and AI is possible across industries. This means that you can find similar patterns in many sectors. Among the very advanced examples of data analytics and AI are in finance or fast moving consumer goods with the goals of detecting fraud, money laundering or consumer preferences. Across all these application areas, my observations were always similar and confirm the desperate need for standardization and increased data quality. Even large companies that do much data analytics and have large volumes of data, need to organize properly for data analytics and innovation. Organizations that were successful in applying data analytics for mortgage calculation or credit card fraud, have adopted these techniques already many years ago. They often started in the 1990s and grew in maturity over the years.

### In summary, how should organizations organize for AI?

As mentioned, data preparation and data cleansing are among the most critical and also the most difficult tasks, which need to be done before any data analytics can start. Organizations should ensure that domain knowledge is available early on in this process. From our experience on EDI clearing center platforms as well as from ERP and e-commerce systems, we know about the critical role of data standardization and semantic interoperability. These systems may be considered as data platforms, which provide flexible and fast access to data. Current data platform innovations like the European GAIA-X or IDS, that are interorganizational

and cross-industry in nature, are more challenging since the business purpose is less obvious than with ERP or e-commerce systems. Mechanisms to define meta data and to apply these structures during data preparation will be an important development to watch.

Dear Yao-Hua, thank you for the interview.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

### References

- Alt, R. (2021). Electronic Markets on the next convergence. *Electronic Markets*, 31(1), 1–9. <https://doi.org/10.1007/s12525-021-00471-6>.
- Heijmann, F., Tan, Y.-H., Rukanova, B., & Veenstra, A. (2020). The changing role of customs: Customs aligning with supply chain and information management. *World Customs Journal*, 14(2), 131–142. <https://worldcustomsjournal.org/archive/volume-14-number-2-september-2020/>.
- Henningsson, S., Gal, U., Bjørn-Andersen, N., & Tan, Y.-H. (2011). The next generation information infrastructure for international trade. *Journal of Theoretical and Applied Electronic Commerce Research*, 6(1), 1–15. <https://doi.org/10.4067/S0718-18762011000100002>.
- Park, T. (2018). *Blockchain is about to revolutionize the shipping industry*. Bloomberg. April 18, 2018. <https://www.bloombergquint.com/technology/drowning-in-a-sea-of-paper-world-s-biggest-ships-look-for-a-way-out>. Accessed 31 July 2021.
- Rukanova, B., Tan, Y.-H., Slegt, M., Molenhuis, M., Rijnsoever, B. V., Migeotte, J., Labare, M. L. M., Plecko, K., Caglayan, B., Shorten, G., van der Meij, O., & Post, S. (2021). Identifying the value of data analytics in the context of government supervision: Insights from the customs domain. *Government Information Quarterly*, 38(1), 101496. <https://doi.org/10.1016/j.giq.2020.101496>.
- Segers, L., Ubacht, J., Tan, Y.-H., & Rukanova, B. (2019). The use of a blockchain-based smart import declaration to reduce the need for manual cross-validation by customs authorities. *Proceedings of the 20th Annual International Conference on Digital Government Research*, 196–203. <https://doi.org/10.1145/3325112.3325264>.
- Stijn, E. V., Rukanova, B., Wensley, A. & Tan, Y.-H. (2009). Moving an eInnovation from a living lab to the real world: Politically savvy framing in ITAIDE's beer living lab. *Proceedings of the Bled eConference*. <https://aisel.aisnet.org/bled2009/31>.
- Tan, Y.-H., Bjørn-Andersen, N., Klein, S., & Rukanova, B. (Eds.) (2011). *Accelerating global supply chains with IT-innovation - ITAIDE tools and methods*. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-15669-4>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.