IoT and Data Challenges

Peter Wittenburg & Leif Laaksonen
June 2017

This is a report about the workshop "Globally Interoperable IoT Identification and Data Processing" held at 6.6.2017 at the IoT Week in Geneva (https://sites.grenadine.co/sites/iot/en/iot-week-2017/schedule/547/IoT+and+Big+Data). The agenda can be found in the appendix. We extracted a few key messages for every speaker, summerised the panel and based on that draw some conclusions.

1. Conclusions

The overarching theme in the discussions was how we can stimulate data exchange (science), data trading (industry) and thus re-use in different contexts. Two sub-topics seem to be most crucial:

- How do data creation, organisation and management be changed in a way to make data intensive projects in science and/or industry much more efficient and thus enable re-use effectively? What is the right level for global intereoperability and which steps could stimulate re-thinking? Is the Digital Object Architecture which puts global Persistent Identifiers (PIDs) that are resolved to useful property information in the core a good kick-off point?
- How can an open data forum be created where data is offered to foster exchange/trading and how does it need to be structured? How can data collectors in particular in industry be convinced to respect the rights of all stakeholders and based on clarified rights became willing to trade data on such a forum.

This report is not the place to give answers to these questions, but the Research Data Alliance (RDA) plenary P11 in Berlin in March 2018 and its side meetings could be a good place to address the questions in more detail.

2. Main Messages

Peter Wittenburg (introductory talk)

- our current data practices are highly inefficient, too costly, not scalable and excluding many
- IoT with its many smart devices all creating data streams requires a fundamental change
- RDA is working on removing barriers at global lelvel to make data work more efficient and produced already a number of relevant results
- globally resolvable persistent identifiers are in the core of new concepts such as Global Digital Objet Cloud and Type-Triggered Automatic Processing which may give directions of how to overcome the huge fragmentation
- RDA is ready to make the bridge towards industrial data challenges

Bob Kahn (keynote talk)

- Internet was successful due to its simplicity, openess, device independency and scalabilty
- Internet on purpose did not address higher level issues such as what to do with the data that is being exchanged via its protocols
- CNRI[1] designed the Digital Object Architecture (DOA) to deal with digital objects on top of the Internet without dealing with the underlying infrastructures and technologies

[1] Corporation for National Research Initiatives

- It would be "simple" again since PIDs that can be resolved to relevant state information are the anchors supported by a common PID access protocol
- the DOA is a perfect answer to the complexity of the IoT data domain

Tobias Weigel

- the international climate modeling community is preparing its new model phase as a preparation for the next climate report
- the CMIP6[2] data model will be core for interoperability of the 100-250 PB of data
- PIDs will become a central role and will be generated for all digital objects and associated with standardised properties which are relevant for efficient data management within their global federation
- PIDs and standardised data types will also be the basis for automatic processing which is a requirement for the future
- the systematic virtualisation as being suggested in RDA's Global Digital Object Cloud is a promissing way to let the user simply operate with Digital Objects (DO) and abstract away from all infrastructure and implementation details.

Tian Ye

- Persistent identifiers (e.g. Handle, Ecode, OID, CSFT, etc.) are important for proper data management in many sectors in China and a few systems are in use already for different application scenarios
- some examples indicate the massive use of PIDs in Chinese IoT industry, such as product traceability, supply chain management, smart agriculture, industrial Internet, and so on.
- CNIC has built the NIOT platform on top of the various independently operating PID services, thus accepts the variety which has emerged and relies on web services (the current major demands) to be able to resolve PIDs
- NIOT is a unique ID platform in China. It provides an application layer routing and addressing solution that establishes a mapping relationship between IoT object ID and its data ID. So, in the network layer, NIOT can adopt TCP/IP architecture or any other network layer routing protocol.
- with NIOT China has realised a national identification service platform that must cope with the various demands put forward by IoT and other industrial fields
- the service platform is based on an increasing number of nodes to support high resolution demands

Peggy Irelan

- data can be seen as the new oil, but it is yet not traded at all - a Marketplace platform for data is urgently required
- a separation of the collection and the transaction process is the basis for more trading
- a few other measures such as a safe place for data, powerful search engines and flexible analytics services are required as well
- standardisation is relevant, however there are many initiatives that deal with different levels of interoperability
- ways for data processing at the earliest point (on the chip) need to be explored as technical innovation must be paralleled with business model innovation

André Zwanziger

- for T-Systems[3] the development of a Multi-IoT Service platform is an urgent need, since it offers amongst others the integration of data from many different sources

[2] https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6
[3] https://www.t-systems.com/de/en

- T-Systems currently is focussing on an all-in-one solution to satisfy the needs of their customers, i.e. they use internal identifiers for the digital objects in addition to the identifiers, type descriptors etc. that identify for example the sensors that are creating the data
- their system would be flexible enough to work with global name spaces
- their Multi-IoT Service platform makes use of the most optimal features from different cloud solution providers which also is a reaction on the customer wishes to not create dependencies
- they see the need of a data market place to improve trading and seem to be ready to react on new market needs

Alexander Ntoko

- for ITU[4] combating counterfeiting in its various forms is one of the big challenges because its solution requires addressing different issues such as trust and interoperability which are considered important for IoT.
- at each step in a production/delivery chain, provenance information including precise identification is essential to trace what happened and using the built-in Public Key Infrastructure (PKI) to establish trust throughout the process.
- steps in a supply chain will all come with their specific identifiers making interoperability a vital element for a solution.
- identifiers need to be associated with the properties of the object (including the unique fingerprint of the object) to check for authenticity.
- ITU-T X.1255 provides the open architecture framework for trust and intereoperabily for IoT identification
- the Handle System has the required features for implementing interoperabilty and trust in IoT.

Jürgen Heiles

- identification is of course required in any system of interacting components, i.e. identification in IoT is of high relevance
- different entities are involved in IoT systems such as the "thing", its virtual entity, an IoT device, an IoT service and finally also a user - they all need to be identified to make interaction between them happen[5]
- The thing in IoT is the object in the interst of the user. IoT devices like sensors and actuators are used to interact with the Thing, but they are not the Thing except for the case the user is interested in the IoT device itself (e.g. supervision and maintenance of the IoT device).
- identifiers of different entities in an IoT system are associated with different requirements and there are practices which cannot be ignored
- communication identifers like IP or Ethernet MAC addresses are an integral part of each communication protocol. They are essential to setup the communication in IoT, but they should not be used as identifiers for things as communication addresses and interfaces of a thing may change.
- AIOTI[6] has a special task force to discuss IoT identifier issues in depth and a survey has been done in the wider IoT standardization, research and user community in order to stimulated the discussion

Juanjo Hierro

- smart solution finding requires gathering and managing context information at large scale

[4] International Telecom Union
[5] Jürgen did not go about identifiers for data being generated by devices.
[6] https://www.aioti.eu

- FIWARE[7] offers a simple standard API to manage context information and supports many IoT protocols
- FIWARE is working out a number of data models for different application domains which are at this moment at a high abstraction level defining terminology
- FIWARE meets the requirements of the Industrial Data Space initiative which is a German initiative to define and organise an industrial data space which can promote data trading
- FIWARE is busy to develop a first open source implementation of this Industrial Data Space

Panel

Here we list some of the major statements made by the panelists, often questions describe best what was said.

- for the huge IoT domain we obviously need new types of search engines to find useful data
- semantic mapping costs a lot of time and in interface design semantic specification is often neglected
- we need to work on a "data market" and need to understand how we can structure it and which role brokers for example can take
- a big question is how we can make data valuable so that data trading becomes popular and a related question is whether academia can help by testing methods
- obviously complexity of systems for IoT data become so complex that only plug&play components will help to not collapse
- of greatest relevance is to create a basic interoperability layer which will synchronise efforts
- need to accept the basic and fundamental role of globally resolvable identifiers for a functioning data universe
- today IP provides world wide interoperability at the network layer[8], data interoperability is especially required at the semantic layer as one of the complex dimensions in the data domain, finding agreements at the semantic layer is not trivial - but they would ease world wide data exchange also across different application domains enormously
- the usage of registered types which are also associated with an identifier will be essential to improve automation
- the blockchain model is a way to structure data so that provenance is inherently included
- how can we ensure that the rights of data owners[9] (who is the owner in case of wearables) is taken care of, will a complete identification of all participating actors and its store in the PID record for example be sufficient to negotiate rights in an automatic scenario
- the lack of quality in data and metadata is a huge problem, i.e. much effort is put in cleansing before data can be re-used, is there a better way to improve quality
- how to decide about the correctness and authenticity of data before re-usage
- metadata is obviously crucial to deterine whether data can be re-used in different contexts, how much context information is required to decide about usability
- in IoT for all steps in a transaction chain validation mesures are crucial to make data valuable
- security at all steps is a major concern in IoT chains to establish trust in the created data, different levels of security can be applied dependent on the application

[7] https://www.fiware.org/

[8] Using the hourglass model for the Internet architecture with IP in the narrow neck point is widely understood. Using this analogy for the data world with the PID in the narrow neck point may lead to wrong conclusions, since the data domain is much more complex.

[9] It was also questioned who the owner in different scenarios is and whether the Facebook model for example is correct.

Workshop at the IOT Week
about

Globally Interoperable IoT Identification and Data Processing

6. June 2017, 9.00-17.00
International Conference Centre of Geneva
www.iot-week.eu

Whether in industry or science, data intensive projects are not at all efficient and cost too many resources excluding many smart people from participation in exploiting the treasure in data. With the current methods of managing and accessing data we are not fit to handle the continuous data streams with high granularity produced by the billions of smart devices being deployed in the coming years. We feel that it is time to change practices radically, but what are promising approaches to overcome the fragmentation?

Our current data practices are not fit to meet the data challenges of the coming decades!

Experts such as Robert Kahn will discuss about concepts that may open the doors for efficient data processing in the future.

Listen to distinguished experts from industry and science and discuss with them about their ideas. We are very much honoured to have one of the inventers of the Internet, Turing Award winner and designer of the Digital Object Architecture, Dr. Robert Kahn, in the group of experts to sharpen our understanding for the challenges and to identify possible solutions. The participation of experts from the worldwide active Research Data Alliance and International Telecom Union will also help to indicate directions.

Agenda

9.00	Peter Wittenburg (RDA)	Welcome and RDA Concepts for efficient data practices
9.30	Robert Kahn (CNRI)	From IP addresses to Persistent Identifiers – principles of Interlinking
10.15		Discussion
10.30	Coffee Break	
	Session 1 – Towards Interoperable Solutions	
11.00	Tobias Weigel (WDCC)	Infrastructure for the Global Climate Reporting

11.30	Ye Tian (CNIC-CAS)	Application of Digital Object Architecture in China
12.00	Peggy Irelan (Intel)	IoT Standards drive Frictionless Analytics
12.30	André Zwanziger (T-Systems)	MISP - Multi-IoT-Service-Plattform
13.00	Lunch Break	
	Session 2 – Towards Interoperable Solutions	
14.00	Alex Ntoko (ITU)	State of discussions in ITU
14.30	Jürgen Heiles (Siemens)	The role of Identifiers in IoT
15.00	Juanjo Hierro (FIWARE)	Supporting Researchers' Data Interoperability: FIWARE for the development of the IoT Cloud for RDA Europe
15.30	Coffee Break	
16.00	Leif Laaksonen (moderator): Robert Kahn, Tobias Weigel, Ye Tian, Peggy Irelan, André Zwanziger, Alex Ntoko, Jürgen Heiles, Juanjo Hierro, Peter Wittenburg	Expert Panel on Globally Interoperable IoT Identification and Data Processing – which ways to go?
17.00	End	

Panel Questions

Panel Objective: the aim of the panel discussion is to explore how industry and science can effectively collaborate on Globally Interoperable IoT Identification and Data Processing and what role RDA can play in facilitating this. At the end of the panel we would like to have a set of concrete actions that can be taken forward in the short-term future.

As a premise to the panel discussion, the following facts about data management and access are important to remember:

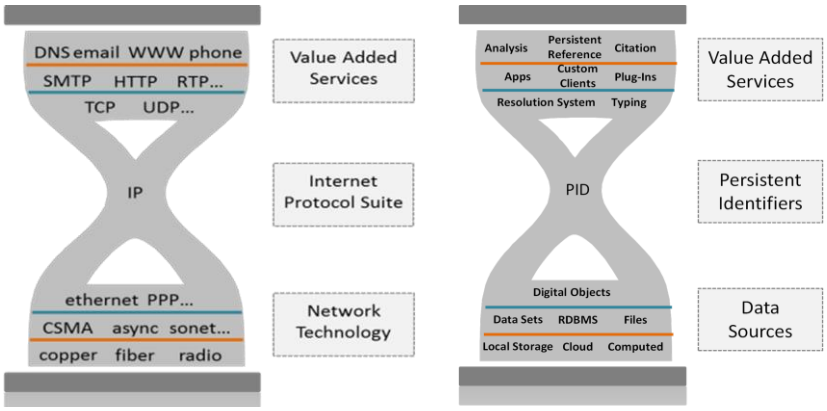
- about 60% of the efforts in data intensive projects in science and industry is wasted on data integration
- about 80% of scientists' time is wasted with simple data management tasks
- about 80% of the data generated is no longer available after a short period of time
- in science and industry we have a large number of infrastructure initiatives and cloud solutions serving urgent needs, but also creating an enormous fragmentation making data integration from different sources an extremely expensive task

According to Intel and Oracle estimates, we will have 50 billion devices all creating continuous streams of highly granular data - orders of magnitude higher in terms of data volumes and complexity than what we currently have. We can simply state that we are not ready to efficiently deal with such data monsters and that we urgently need to change data practices and find global solutions.

This can lead us to a series of questions for the panel:

- For many of us it is evident that we need a change, but which kind of integration solutions do we currently see emerging and what is their potential?

- Expensive data integration from different sources forms a business model today. Is this model sustainable? How much will it dampen IoT business over time?
- What would be the gains compared to the losses for industry to adopt global interoperability standards?
- How far can we go with global agreement forming? Do we need to stick with an hourglass model as in the case of Internet (see the diagram in the attachment)? Is the hourglass still the best model?
- What is the role of the academia in globally interoperable IoT identification and data processing?



Speakers

Jürgen Heiles

Juergen Heiles has over 30 years of experience in Information and Communication Technologies from Satellite Communication to Optical Networking, IPTV, Smart Grid, IoT and Industrial Internet. He has been involved in standardization since the late 80s at ETSI, ITU, Open IPTV Forum, European Smart Grid Coordination Group and various other groups. Currently he oversees the Digitalization related standardization activities at Siemens Corporate Technologies. He is and co-leader of the IoT Identifier task group in the Alliance for Internet of Things Innovations (AIOTI) and a contributor to IEEE P2413 Architecture Framework for Internet of Things. He graduated from the University of Applied Sciences Koblenz, Germany.

André Zwanziger

André Zwanziger is Enterprise Architect and member of the Lead IoT Architects Team at Digital Division/T-Systems International GmbH (http://www.t-systems.com). T-Systems is a subsidiary of Deutsche Telekom AG with main focus on large enterprise customers. The Lead IoT Architects Team invented the award-winning Multi-IoT-Service-Platform in 2015/16 that is the fundamental base for IoT products of Deutsche Telekom and T-Systems (Experton I4.0/IoT Leader 2017 - http://bit.ly/2oLJfZM). Dr. André Zwanziger is author of articles in the

topics of DC automation, performance measurement for service management and IoT as well as a contributor for oneM2M (http://www.onem2m.org/). He is also part of the IoT strategy team at Deutsche Telekom.

Peggy Irelan (https://www.linkedin.com/in/peggy-irelan-4138091)

Peggy is an Intel Fellow, a Solution and Data architect, in the Data Center Solutions Group (DSG). Peggy provides technical leadership to the strategic DSG growth projects as well as lead the definition of data architecture, data models and technical governance model for DSG. Previous, Peggy was in Internet of Things Group working as a technology strategist and delivering end-to-end IoT analytics products in Intel® Decision Suite. This including sensors, software, and algorithms. These products turned the customers' data into trusted actionable intelligence to optimize their business. Previous to that, she was the chief architect in SSG, of Intel's Architecture for Monitoring Initiative. She led the initiative to create an innovative, reliable, and easy to use IA software and hardware platform monitoring and telemetry technology across all Intel product lines. This included the software framework solution that enables monitoring and telemetry driven Software & Services. Peggy started her career at Motorola where she developed and integrated MEMS sensor control systems for everything from cell phones, to run-flat tires, to air bags. Peggy has a B.S. in Computer and Electrical Engineering from Purdue University.

Juanjo Hierro (https://www.linkedin.com/in/jhierro)

Currently Chief Technology Officer (CTO) of the FIWARE Foundation. Also chairman of the FIWARE Technical Steering Committee. FIWARE is an Open Source initiative whose mission is to build an open sustainable ecosystem around public, royalty-free and implementation-driven software platform standards that ease the development of new Smart Applications in multiple sectors. The FIWARE platform combines components enabling the connection to the Internet of Things with Context Information Management and Big Data services on the Cloud. More info about FIWARE can be found at its website (http://fiware.org), YouTube channel (http://www.youtube.com/user/FIWARE) or following @FIWARE on Twitter.

Robert Kahn (https://en.wikipedia.org/wiki/Bob_Kahn)

Robert E. Kahn is Chairman, CEO and President of the Corporation for National Research Initiatives (CNRI), which he founded in 1986 after a thirteen year term at the U.S. Defense Advanced Research Projects Agency (DARPA). CNRI was created as a not-for-profit organization to provide leadership and funding for research and development of the National Information Infrastructure. Among many other awards, Bob Kahn was awarded the SIGCOMM Award in 1993 for "visionary technical contributions and leadership in the development of information systems technology", and shared the 2004 Turing Award with Vint Cerf, for "pioneering work on internetworking, including the Internet's basic communications protocols and for inspired leadership in networking."

Leif Laaksonen (http://www.leiflaaksonen.eu)

Leif Laaksonen is currently a Director at CSC, the Finnish IT Center for Science located in Espoo, Finland, with responsibilities in international collaboration and project coordination/ management. His education is in quantum chemistry with a flavour of bioinformatics, from the late 80-ties and early 90-ties through practice. He has a long tradition in European research collaboration through chairing the e-Infrastructure Reflection Group (e-IRG) for the period 2007 – 2008 and 2009 – 2010 and serving as the e-IRG representative in the European Strategy Forum on Research Infrastructures (ESFRI) and as a e-IRG member in the ESFRI initiated Working Group on investment strategies in e-infrastructures.

AlexanderNtoko (https://www.itu.int/en/ITU-T/Workshops-and-Seminars/20160628/Pages/NTOKO-Alexander.aspx)

Alexander NTOKO is the Chief of the Operations and Planning Department in ITU where he plays a critical role in crafting and executing the strategic movements of ITU's Telecommunication Standardization Sector (ITU-T). He coordinates ITU activities on the Digital Object Architecture (DOA) and represents ITU in ICANN's Governmental Advisory Committee (GAC). Prior to taking up his current position, he was Head of ITU Corporate Strategy Division (CSD) where his responsibilities included ensuring organization-wide consensus and support for the development, implementation and evaluation of the ITU Strategic Plan. He assisted and advised the ITU Secretary-General in policies and strategies and was responsible for the coordination of ITU activities in many areas including Internet and cybersecurity.
From the early 1990s, he played key role in the introduction of Internet and Cybersecurity to ITU. Since 1998, he has managed the implementation of projects on Internet, ICT applications (e.g., e-health, e-government, e-agriculture, e-education, e-payment and e-business) and cyber security including the use of advanced security technologies (biometric authentication and Public Key Infrastructure) in developing countries from all regions of the world. He obtained Bachelors (BSc) and Master of Science (MSc) degrees in Computer Science from the State University of New York.

Tobias Weigel (https://www.dkrz.de/about/Organisation/mitarbeiter/TobiasWeigel)

Tobias Weigel is a computer scientist working at the German Climate Computing Center (DKRZ) particularly in the areas of data infrastructures and persistent identifier services. Activities he is involved in include EUDAT services as well as big data management and analytics services for the Coupled Model Intercomparison Project phase 6 (CMIP6), which will generate the largest assembly of climate modelling data over the next years. Tobias holds a PhD in computer science from Hamburg University and is editorial board member of the CODATA Data Science Journal. Tobias was and is co-chair of

multiple working groups of the Research Data Alliance focused on PIDs, typing and actionable collections, and is a member of its Technical Advisory Board.

Peter Wittenburg (http://www.mpcdf.mpg.de/~pewi)

Peter Wittenburg was for many years technical director at the Max Planck Institute for Psycholinguistics (http://www.mpi.nl) and his group was responsible for technology and methodology development to facilitate the understanding how human brains process and acquire languages. Due to the outstanding work of his group he was appointed to become a member of the IT advisory board of the Max Planck Society. From about 2000 his group had leading roles in infrastructure building at national, European and international level (DOBES, CLARIN, EUDAT, ISOCAT). Due to his experience with the huge fragmentation he was one of the founders of the Research Data Alliance (http://rd-alliance.org). Recently he was invited by the German minister for education and science to lead a working group in the realm of the German IT Summit (https://www.bildung-forschung.digital/de/plattform-digitalisierung-in-bildung-und-wissenschaft-1717.html) and established intensive interactions with German IT industry.

Ye Tian (http://www.escience.cn/people/tianyeiot/index.html)

Tian is the executive deputy director of the Internet of Things Information Technology and Application Laboratory of the Computer Network Information Center (CNIC), Chinese Academy of Sciences (CAS). He is also the senior member of the China Computer Federation (CCF) and the liaison of the CCF Task Force on Big Data. Dr.Tian is responsible for the national information projects to ensure that every project can be carried out and be accomplished on time. Those projects include the National IoT Industrialization project, the China's Next Generation Internet project, The National Key Technology R&D Program, and so on. As the executive manager of the China National IoT Name Service Platform, he takes charge of the whole construction of the technology development and the nodes deployment.