# Myriads of data, myriads of devices: self-awareness of the Ad-hoc

**Ioannis Chatzigiannakis, Othon Michail, Georgios Mylonas, and Paul Spirakis**

*Designing highly distributed dependable systems offers a means of dealing with open data.*

Recent advances in pervasive computing and the novel Internet technologies coming out of research labs provide a glimpse of the future, including smart cities, streets, schools and much more. By using the Internet to connect real-world sensors and control mechanisms to cloud-based services that pull streams from other data sources, we create an opportunity for new wide-impact services and products. Future systems will orchestrate myriads of units/nodes, Web services, business processes, people, companies and institutions. These will be continuously integrated and connected, while preserving their individual properties, objectives and action.

We are already experiencing a new stage of Web evolution. New data-centric applications and services are emerging, the value of which is based on using open data. Retailers use open data to combine radio frequency identification, near-field communication or sensor data, social media data and GPS coordinates to evaluate location, product selection and individual profiles to deliver highly personalized services. For example, the UK government[1] leveraged crowdsourcing to improve its data set of countrywide bus-stop locations when the locations of 18,000 bus stops turned out to be inaccurate.

According to IBM,[2] we create 2.5 quintillion bytes of data every day. Currently, 90% of the data available has been created in the past two years, and the amount of data is expected to increase exponentially. This data is expanding rapidly as enterprises capture more information in greater detail (see Figure 1). Multimedia is becoming more common, and social media conversations continue to grow in popularity. In general, the Internet is more pervasive than ever. Open data provides tremendous benefits that are already established in a number of sectors in government and business.

Future ICT systems will have to continuously monitor vast amounts of data, and classify and deliver it in real time to the
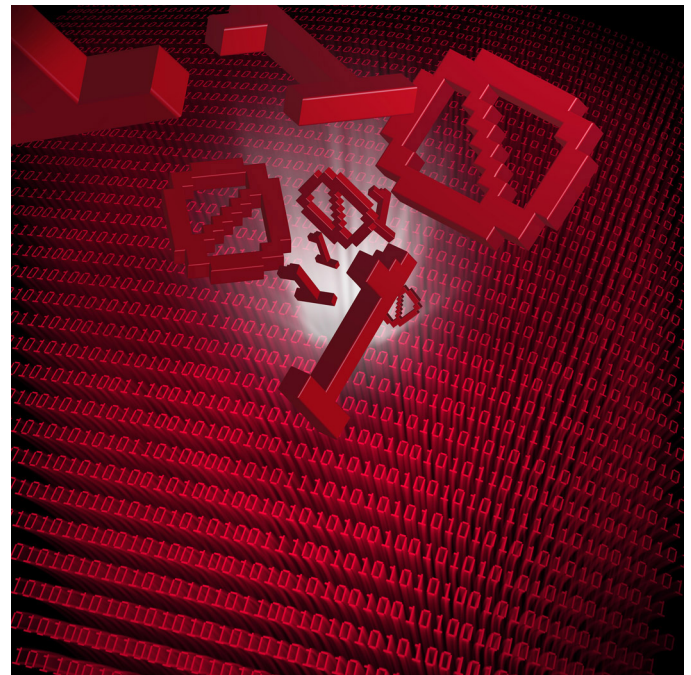


**Figure 1.** *Today, enterprises capture more data, and in more detail, resulting in exponential growth in the amount of information needing to be stored and processed. (©Blueximages | Dreamstime.com)*

appropriate users. The performance of the next generation of services will primarily be characterized by the volume (amount), the velocity (speed of creation and modification), and the variety (types and sources of structured and unstructured data) of linked open data these services use. It is evident that processing and managing several quintillion bytes of data on a daily basis is beyond the capabilities of any existing centralized processing system.

Recently, the EU-funded Future and Emerging Technologies project FRONTS[3] proposed decomposing processing into very simple computational entities that examine tiny fractions of open data. These computational entities can be hosted by the

myriads of resource-limited pervasive devices that are embedded into environments, appliances and everyday objects. In this way, they form a large population of tiny agents that examine and interact ubiquitously, hidden in the fabric of everyday life. Interestingly, populations of extremely limited computational entities, if organized appropriately, can form powerful systems[4] that collectively process and evaluate the properties of structured open data.

As an example, let's consider mining of real-time social media data to extract business intelligence data elements. Current solutions rely on centralized monitoring, processing and integration with enterprise IT platforms. In contrast, FRONTS programs tiny agents to inspect information using simple local rules. It then deploys them as plug-ins to browsers and as tiny processes for smartphones. These agents interact by exchanging short messages. This is a continuous, never-ending process that eventually reaches a state in which the agents develop a self-understanding of the global state.

Consider an enterprise that wishes to characterize its customers' opinion on a particular product line by monitoring their comments. Here, no single comment is sufficient to immediately reach a decision. Instead, we face a sparse set of comments that identify characteristics of a fad. Each tiny agent stores only two bits of information: P if it has encountered at least one positive comment, N if it has encountered at least one negative comment and U if it is still undecided. Initially, all agents start at state U. An agent sets the two bits to match its own opinion. If an agent is in state U, and encounters a positive (or negative) comment, it changes to P (or N). If an agent is in state P (or N) and detects a negative (or positive) comment, it changes back to U. If two agents interact and are in the same state, they maintain their states. If one is in P and the other in N, they both change to U. If one is in P (or N) and the other in U, then the undecided agent changes state to P (or N). More formally, the protocol is (P,U) → (P,P), (N,U) → (N,N), (P,N) → (U,U).

One can view this protocol as a propagation of conflicting epidemics: the epidemic of P and the epidemic of N. The agents in state U are not infected, and agents in states P and N attempt to infect the agents they meet with their respective state. Such agents immediately infect a non-infected agent and cure an agent of the opposing epidemic. This simple protocol manages to detect the opinion expressed by the majority: the dominating opinion is eventually propagated to all the agents of the population. Interestingly, the more relevant the comments inspected by each agent, the faster the protocol reaches a safe conclusion.

The example demonstrates how a vast number of local, limited and appropriately organized computations can be used to analyze large volumes of data in a correct and fast way. The FRONTS approach totally avoids central storage of extracted data and, by design, preserves the privacy of individual users.

Yet there is a need for further research beyond FRONTS. These systems can be truly useful only if they are dependable, so that society can trust them and broadly participate in them. Existing methods do not seem to be able to achieve such a vision. Distributed computing and game theory have not managed, as yet, to jointly describe huge collections of autonomous entities that care for global welfare. Also, modern distributed computing has not yet developed a set of tools that are adequate for extreme network dynamicity and wide openness. We strongly believe that future networked systems, as outlined here, will contribute substantially towards unleashing the potential of open data.

## Author Information

**Ioannis Chatzigiannakis, Othon Michail, Georgios Mylonas and Paul Spirakis**
Computer Technology Institute & Press 'Diophantus' (CTI)
Patras, Greece

**References**
1. N. Shadbolt, *Open for business*, **Think Quarterly**, March 2011. http://www.thinkwithgoogle.co.uk/quarterly/data/nigel-shadbolt-open-data.html
2. http://www-01.ibm.com/software/data/bigdata/ Bringing big data to the enterprise. Accessed 25 October 2012.
3. http://fronts.cti.gr Foundations of Adaptive Networked Societies of Tiny Artefacts. Accessed 25 October 2012.
4. I. Chatzigiannakis, O. Michail, and P. Spirakis, *New models for population protocols*, **Synth. Lect. Distrib. Comput. Theory 2** (1), pp. 1–156, 2011. http://www.morganclaypool.com/doi/abs/10.2200/S00328ED1V01Y2011-01DCT006?journalCode=dct