# Folk-IS: Opportunistic Data Services in Least Developed Countries

Nicolas Anciaux, Luc Bouganim, Thierry Delot, Sergio Ilarri, Leïla Kloul, Nathalie Mitton, Philippe Pucheral

# Folk-IS: Opportunistic Data Services in Least Developed Countries

N. Anciaux[1, 2], L. Bouganim[1, 2], T. Delot[3,1], S. Ilarri[4], L. Kloul[2], N. Mitton[1], P. Pucheral[1, 2]

[1] INRIA, France *fname.lname@inria.fr*
[2] PRISM, UVSQ, France *fname.lname@prism.uvsq.fr*
[3] LAMIH, UVHC, France *Thierry.Delot@inria.fr*
[4] Univ. of Zaragoza, Spain *silarri@unizar.es*

## ABSTRACT

According to a wide range of studies, IT should become a key facilitator in establishing primary education, reducing mortality and supporting commercial initiatives in Least Developed Countries (LDCs). The main barrier to the development of IT services in these regions is not only the lack of communication facilities, but also the lack of consistent information systems, security procedures, economic and legal support, as well as political commitment. In this paper, we propose the vision of an infrastructureless data platform well suited for the development of innovative IT services in LDCs. We propose a participatory approach, where each individual implements a small subset of a complete information system thanks to highly secure, portable and low-cost personal devices as well as opportunistic networking, without the need of any form of infrastructure. We review the technical challenges that are specific to this approach.

## 1. INTRODUCTION

As a citizen of a developed country, Alice receives most personal data electronically (e.g., salary forms, banking statements, medical records) and stores them in the cloud where service providers deliver a myriad of digital services in the context of healthcare, education, and business. But what if Alice lives on the opposite side of the digital divide? Least Developed Countries (LDCs) are those countries that meet United Nations (UN) criteria in terms of poverty, human resource weaknesses and economic vulnerability. These countries definitely lack IT infrastructures. Nevertheless, many reports (e.g., [7, 9, 12]) emphasize that IT is called to play a catalytic role in LDCs, helping to establish primary education, reduce mortality and boost individual commercial initiatives.

While 60% of the population in LDCs is already covered by a mobile cellular signal, only 0.5% has a mobile broadband subscription and a 3G service is offered only in at most 25% of the LDCs, very often at a prohibitive cost [9]. Hence, mobile phones in these areas are primarily feature phones used for voice and SMSs, not for data-driven applications. According to many analysts, this situation cannot evolve rapidly due to a combination of technical, economical and organizational barriers [9]. A few pioneering proposals [11, 14] tried to overcome these limitations by mounting mobile access points on vehicles to transport data from one place to another. Google project Loon is a more ambitious attempt to connect people in rural and remote areas and bring people back online after disasters thanks to a network of high altitude balloons. However, it is unclear if such a network infrastructure is durable, since current balloons can only function for a few weeks. For its part, Internet.org promotes low cost wireless handsets, compression of Web pages, and data caching on the edge of the network, to reduce data transfer. These different initiatives demonstrate the growing interest, as well as the high difficulty, to integrate LDCs in global IT networks.

But low connectivity is not the ultimate and unique barrier. Several solutions have considered the use of mobile phones (and text-messages data transfer) to address specific issues like tracking vaccine cold chains [6], improving agriculture [13], or easing administrative procedures [10]. Despite their undisputed interest, these initiatives remain confined to specific applications and their generalization faces many obstacles: lack of global information system infrastructures, maintenance of the system, security concerns, etc. Hence, we consider as characteristics of LDCs not only the lack of communication facilities, but also the lack of consistent information systems, security procedures, economic and legal support, as well as political commitment.

According to Non-Governmental Organizations (NGOs), four main requirements must be met to build a practical technical solution: (1) **privacy protection**: a major prerequisite due to local opaque practices and the lack of any security infrastructure (coercive laws, secured servers, trusted authorities, etc.), leading to a self-enforcement of privacy principles; (2) **immediate personal benefit**: that should be provided to each user because of the lack of strong economical or political incentives to impose the solution; (3) **self-sufficiency**: the solution must not rely on a hypothetic improvement of the existing software and hardware infrastructure; and (4) **very low deployment cost**: the usual scale being a few dollars per user. Besides this, **users' empowerment** is crucial to make the solution sustainable in LDCs, and its maintenance should ideally generate a **source of revenue for new local jobs**.

In [1], we proposed the vision of *trusted cells*, a data platform for personal data services where the shared infrastructure (e.g., the cloud) is untrusted, while personal devices (e.g., smart phones) are trusted execution environments. In this paper, we revisit this vision to the context of LDCs. We propose, *Folk-enabled Information System (Folk-IS),* a new paradigm based on a fully decentralized and participatory approach, where each individual implements a small subset of a complete information system without the need for infrastructure. As trusted cells, Folk-IS builds upon the emergence of highly secure, portable, low-cost storage and computing devices, called hereafter *Smart Tokens*. Here, however, the focus is on the low-cost of ownership, deployment and maintenance, and on the absence of a networked infrastructure. With Folk-IS, and thanks to smart tokens, people will transparently and opportunistically perform data management and networking tasks as they physically move, so that IT services are truly delivered by the crowd.

We do not argue that Folk-IS is the ultimate solution. The future of IT in LDC will probably be multiform, the problem being important and complex enough to leave room for complementary initiatives. Folk-IS has the salient characteristics to enable a


*Proceedings of the VLDB Endowment, Vol. 7, No. 5*

smooth and incremental deployment of an information system in a purely infrastructureless context while taking advantage of existing elements of infrastructure, if any, to improve its own behavior. This paper shows that this paradigm technically makes sense, conforms to the requirements mentioned above, and opens important and exciting research challenges.

## 2. ARCHITECTURE AND SCENARIOS

The main Folk-IS elements are the following ones:

*Smart Tokens*: Smart Tokens appear today in various form factors and may have different hardware characteristics. In the Folk-IS context, Smart Tokens embed at least: *(1)* enough stable storage to host the complete digital environment of its holder, *(2)* enough tamper-resistant computing resources to run a server managing the data and enforcing access control rules, and *(3)* a biometric sensor to authenticate users (e.g., a fingerprint reader, well adapted to illiterate people). Smart Tokens require also input/output capabilities to interact with users and communication facilities (e.g., short-range communications) to exchange messages.

*Shared devices:* To meet low-cost requirements, a Smart Token may inherit part of its functionalities from the terminals it connects to. While storage and security resources cannot be delegated to shared devices without compromising data availability and introducing the risk of class-breaking attacks, the I/O capabilities of shared devices (e.g., screen, keyboard, etc.) could be naturally used. Similarly, shared devices can act as a relay when connected to the Smart Tokens, compensating their lack of communication facilities. Shared devices are either made publicly available in specific places or owned by local workers.

Figure 1 shows concrete examples of Smart Tokens and shared devices (e.g., off-the-shelf PCs and tablets or specific devices containing only validated software and hardware). The *Basic* Smart Token embeds only mandatory resources, namely a Flash stable storage, a Secure microcontroller (SMCU) [4], and a fingerprint reader. It achieves the lowest cost and the highest robustness at the expense of needing a shared device to be used. The left-end side of the figure shows a real product that we used in a field experiment [2], provided by Gemalto for a few dollars. The *Self-powered* Smart Token is a bit more expensive but includes also a speaker and a microphone for basic user interactions, a battery, solar cells, and wireless communication to enable message exchanges without the need of shared devices.

*Folk-IS Personal Node (Folk-node)*: a Folk-node is associated with each individual and refers to the combination of a Smart Token and the embedded software components required to manage the holder's data, act as a network node, and enforce the security of the whole system. Based on Folk-nodes, we can set up an ad hoc delay-tolerant-network [5], called in the following *Folk-enabled Network (Folk-Net),* enabling the transfer of messages among Folk-nodes, and from Folk-nodes to Internet access points, without requiring any existing communication infrastructure. In Folk-Net, communications are opportunistic by nature, and follow a carry-and-forward protocol.

*Folk-enabled Information System (Folk-IS)*: based on the existence of Folk-nodes, we can devise a system implementing the three main functions of an information system as follows:

- *Communication management:* by carrying and routing data, each Folk-node acts as an active node in the Folk-Net. Assuming that each Folk-node maintains a history of its moves (e.g., by gathering the GPS coordinates of all the devices it connects to), it can forward messages to encountered Folk-nodes whose moving profile best matches the recipient's location, thus providing a much better resource utilization of the Folk-Net than a basic flooding protocol.

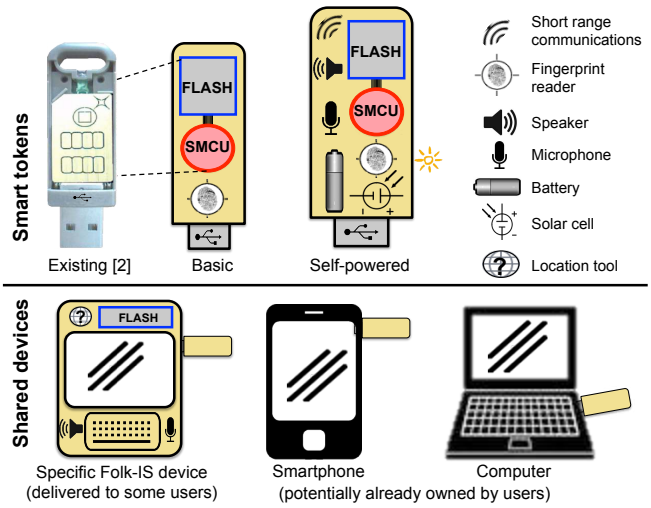


**Figure 1. Smart Tokens and shared devices**

- *Data management:* each individual centralizes in his Folk-node all his personal data (e.g., medical records, administrative documents, credentials). Each time the holder interacts with a data source (e.g., by physically meeting a data provider, like a doctor, or by receiving a document through the Folk-Net), the application simply inserts these data in a local Folk-node database. From the information system viewpoint, the global database is the "union" of all these local databases. Each Folk-node is assumed to participate in common services, like processing global queries (broadcasted through the Folk-Net) and ensuring data durability (thanks to data replication among Folk-nodes).
- *Access control & authentication:* each Folk-node controls the access to the data it hosts and strongly authenticates the requesters. Performing this control on the user's side is the ultimate solution to increase the holder's confidence, enforce his consent, and minimize the benefit/cost ratio of an attack. Indeed, the complexity of attacking the system is increased by the Smart Token tamper-resistance and by the obligation to be physically in contact with it to attack it. In parallel, the benefit of the attack is limited to disclosing/corrupting the data of a single individual. Note that the holder himself must authenticate and does not have all the privileges on his own Folk-node (e.g., he cannot tamper his own medical data to prescribe himself new drugs). In addition, messages carried by a Smart Token or replicas from other individuals stored locally are encrypted with keys never available to that Smart Token.

Hence, the complete system (data storage and network facilities) progressively deploys itself as Folk-nodes are distributed to the participants. Folk-IS is by construction highly redundant and robust, it does not require any central administration, and its global cost is proportional to the scale of the targeted population. As discussed next, some Folk-IS functionalities could be delegated to specific workers (e.g., people renting terminals or acting as postmen) in order to improve the quality of service while creating a new local economic model (e.g., like people renting their cell phones in LDCs [3]). If the infrastructure is partially available, the Folk-IS quality of service also improves accordingly, decreasing the network latency by shortening the route to the nearest Internet access point.

Figure 2 illustrates the Folk-IS mode of operation. It shows two rural communities, residents (black icons with letters) and their possible moves (grey icons), a *school* and a *first aid* room used by

residents from both communities, and an Internet access point. Data exchanges through terminals are represented by dashed lines (e.g., the Folk-node of person A is used from the terminal of nurse N in the infirmary), and short-range data exchanges are represented by pink halos (e.g., between A and B in *Community 2*). E serves as a *netman* (i.e., a postman conveying digital messages) and carries data when travelling from place to place (e.g., the school, the infirmary, and the Internet access point) thanks to his Folk-node. Different routing paths to transmit a message from A to the Internet are represented, e.g., the path ABTE: A → B → T (teacher) →E (netman) → Internet. These paths benefit from the physical mobility of people and their devices.

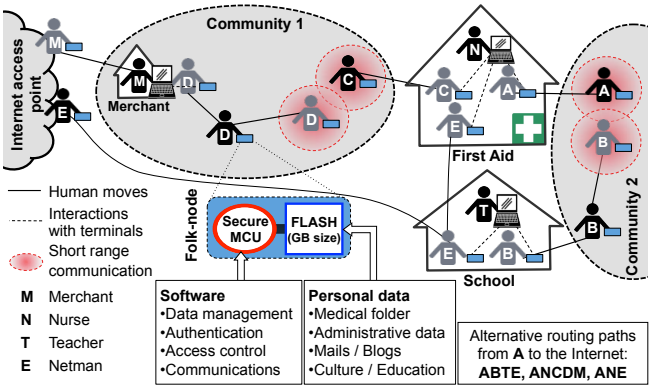


**Figure 2. Folk-IS: mode of operation**

Based on this mode of operation, several concrete scenarios can be envisioned. For illustration, we show an example below.

*Healthcare scenario:* In most LDCs, healthcare is provided by local nurses/doctors working in first aid rooms or intervening in rural communities during drought periods or health programs. They only possess very basic medicines and equipment. Needless to say, they do not benefit from Electronic Health Records, which are already highly difficult to organize in developed countries. Even paper-based medical records are too complex to maintain because many people have no identity document to link them to their records and people move according to seasons and drought periods. By using a Folk-node, each patient owns his complete and up-to-date medical folder. Without any Internet connection, a nurse can access the fingerprint-authenticated patient's medical folder, and append diagnosis and treatment information. She may request medical advice from a distant clinic, by transmitting documents (e.g., a picture of the patient's injury) from the patient's Folk-node through the Folk-Net. She can notify the patient a few days later, still through the Folk-Net, in case a serious problem has been detected. Global queries can also be broadcasted through the Folk-Net to conduct epidemiological studies or maintain health indicators on populations unreachable so far. Hence, each Folk-node acts as a smart Personally Controlled Electronic Health Records.

*Generalization:* Many usual scenarios can be transposed to rural communities thanks to Folk-IS, provided that they accommodate high-latency data exchanges: scholar folders for children, e-administration procedures (e.g., drought warnings, recording births and deaths), personal data applications (e.g., email, social or networking services), etc. Folk-IS also paves the way to new practices like sharing pieces of cultural heritage (e.g., traditional songs, beliefs and religion, etc.) with the outside world. This may generate a new economic model where every actor is paid according to a specific Digital Rights Management (DRM) model. Similarly, people playing the role of a *netman* can be paid according to the volume of transported data and distance covered. In both cases, the exact contribution of each actor can be recorded and certified by his tamper-resistant Folk-node.

# 3. DATABASES CHALLENGES

Several challenges of Folk-IS are at the crossroad of Computer Science, Economics and Social Science. Issues linked to economic models are crucial for the adoption of the system, and the problems linked to the low level of education, lack of training, specific beliefs and norms of potential users need to be addressed. In the following, we focus solely on database challenges.

**(1) Co-design of the embedded data management engine.** A primary role of a Folk-node is to ensure secure storage and sharing of all data forming the holder's digital environment: user authentication, secure data storage, query evaluation, and access control enforcement. Designing these components is very challenging due to the inherent hardware constraints of smart tokens, namely the tiny RAM of the secure microcontroller, the NAND flash memory that badly supports random writes, and the need to protect the confidentiality and integrity of the stored data. These constraints lead to contradictory objectives: executing queries with a tiny RAM entails indexing massively the embedded database, while index updates generate fine-grained random writes, and then unacceptable NAND flash write costs and cryptographic overhead. In addition, small hardware variations (i.e., varying the precise characteristics of each element) may greatly impact their cost, energy consumption, and performance. Existing embedded and lightweight DBMS products target devices far more powerful than smart tokens, and state-of-the-art research solutions do not explicitly tackle these issues. Initial work has been done to handle classical smart token constraints [2], but the challenge here is to co-design and implement the best storage, indexing and query engine to reduce the overall cost of the device and its energy consumption, and provide a high level of flexibility.

**(2) Folk-Net routing and mobility prediction protocols.** The Folk-Net routing protocol aims at delivering the data exchanged among Folk-nodes, workers' devices, and remote Web servers. Given the lack of infrastructure and the highly dynamic nature of Folk-Net, the routing protocol has to rely on opportunistic communications. Thus, the choice of the best candidate(s) to carry the encrypted data towards their destination is a primary concern. Another concern is the energy limitations, which have to be taken into account in the design of the routing scheme. Geographic routing protocols are interesting candidates to avoid flooding the network, as they do not need to maintain routing tables and work nearly stateless. However, existing georouting protocols can hardly be considered because they rely on the exact locations of individual nodes. In our context, the location of the device must be approximated from interactions with a small subset of localized Folk-nodes, shared devices, or fixed nodes. New routing protocols mixing geolocation approximation, mobility prediction strategies, and social interactions (e.g., [8]), definitely deserve to be studied to select the best "data mules".

**(3) Application, data model and access control deployment.** Folk-IS enables many important applications, like healthcare and e-administration, some of them novel by their purpose (e.g., sharing pieces of cultural heritage) or by their target (e.g., epidemiological studies on populations unreachable so far). The infrastructureless context introduces new challenges with respect to application deployment, unified data modeling, identity verification, access control, and user's consent. Typically, data standardization, central application stores, or central authorities

identifying people, delivering certificates and enforcing access control rules, cannot be assumed. Conversely, a salient feature of Folk-IS is the ability to push the control at the edge of the network, that is to say (1) within each tamper-resistant Smart Token, and (2) through face-to-face interactions between users (a de-facto user's consent). This may enable the reestablishment of local spheres of trust (e.g., NGO producing applications, data models, home-made identification information, access control policies, and pushing them at the Folk-node level through Folk-Net). This paves the way to a semi-decentralized way of managing and deploying applications, each Folk-node guaranteeing that (i) data produced by a given organization will not leak outside that organization, and (ii) the data owners' privacy is always respected.

**(4) Evaluation of global queries on a population of Folk-nodes.** Traditional techniques used to process queries in distributed databases or in peer-to-peer networks are not suitable, as they assume a good knowledge about the data location in the network. Here, data will be dynamically distributed over a network of nodes that may be accessible or not at a certain moment, and the carrier of a replica of a data item may change at any time. Moreover, we cannot assume that all data relevant to a given query will always be available and retrievable in a reasonable time period. So, we should assume the possibility of approximate answers, decide how to identify the relevant data sources without overloading the network (e.g., based on spatial conditions), how to route queries and results to their recipients (while preserving autonomy) using the physical mobility of nodes, and how to determine when a query and its associated routing tasks have to be finished. Moreover, new types of queries could be of interest in this context, requiring new query processing techniques; for example, reachability queries [15] could be used to study the possibility of propagation of a disease by analyzing past trajectories.

**(5) Structuration and calibration of the Folk-IS architecture.** Folk-IS is built on a large number of highly-secure but seldom-available Folk-nodes (basic or self-powered) on the one side, and on less secure but more accessible and powerful shared devices on the other side. This unusual asymmetric architecture requires deeply rethinking the overall organization of an information system. This means distributing software resources on the hardware elements of the architecture, such that local and global applications can run and provide results with acceptable performance, security and resiliency. This also means associating different responsibilities to different devices (e.g., *super-nodes*), specific roles to humans (e.g., *netmen* to reduce latency), and designing new distributed protocols for global functionalities like query evaluation or data durability. For the latter, data replication is required because Folk-nodes may be lost, stolen, broken, etc. Data could be replicated based on the user mobility profiles. Confidentiality of replicas can be achieved through encryption, leading to the problem of choosing the most adequate set of Folk-nodes to hold a copy of the keys or of key shares (e.g., based on trust or similar mobility profiles). Finally, new simulation models are needed to calibrate IT resources according to the target applications and the local habits of residents. The objective is to determine suitable network topologies (given a certain density of Folk-nodes, communication frequency, etc.) with viable incremental deployments. Those simulation models will also help quantify the expected gain in terms of social and economic sustainability, which is key for their long-term adoption

# 4. CONCLUSION

As mentioned in [3], time has come for research works, not only commercial initiatives, addressing the expectations of 80% of the population living outside developed countries. The *Folk-IS* paradigm combines low-cost secure devices, embedded software components and opportunistic communications, to meet fundamental requirements of LDCs. The promise of the solution is to guarantee high privacy standards at very low cost (a few dollars per user), granting users access to innovative personal services with the ability to benefit from future infrastructure improvements. The maintenance and performance improvement of the system can be a source of empowerment with new local jobs, crucial to make the solution sustainable. We have presented an initial architecture and identified important challenges, which pave the way to exciting future works for our research community.

***Acknowledgment:*** We warmly thank Léo Dayan, scientific director of APREIS NGO and of the Nomadic World University for Sustainable Development, and Pascale Pollack, director of E-NEXUS NGO, for precious discussions and feedback on the IT requirements for LDCs.

# 5. REFERENCES

[1] N. Anciaux, P. Bonnet, L. Bouganim, B. Nguyen, I. S. Popa, P. Pucheral, “Trusted Cells: A Sea Change for Personal Data Services”. CIDR, 2013.

[2] N. Anciaux, L. Bouganim, Y. Guo, P. Pucheral, J.-J. Vandewalle, S. Yin, “Pluggable Personal Data Servers”, ACM SIGMOD, 2010.

[3] E. Brewer, M. Demmer, B. Du, M. Ho, M. Kam, S. Nedevschi, J. Pal, R. Patra, S. Surana, K. Fall. “The Case for Technology in Developing Regions”, Computer 38(6), 2005.

[4] D. Bursky, “Secure Microcontrollers Keep Data Safe”, PRN Engineering Services, http://tinyurl.com/secureMCU, 2012.

[5] Y. Cao, Z. Sun, “Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges”, Communications Surveys & Tutorials, IEEE 15(2), 2013.

[6] R. Chaudhri, G. Borriello,R. J. Anderson, “Monitoring Vaccine Cold Chains in Developing Countries”, IEEE PerCom, 11(3), 2012.

[7] R. Coceres, E. M. Belding, T. S. Parikh, L. Subramanian, “Information and Communication Technologies for Development - Guest Editors' Introduction”, IEEE PerCom, 11(3), 2012.

[8] L. Gao, M. Li, A. Bonti, W. Zhou, and S. Yu, “Multidimensional Routing Protocol in Human-Associated Delay-Tolerant Networks”, IEEE TMC, 12(11), 2013

[9] ITU, “The Role of ICT in Advancing Growth in Least Developed Countries – Trends, Challenges and Opportunities”, 2011.

[10] V. Ndou, “E-Government for Developing Countries: Opportunities and Challenges”, EJISDC volume 18, 2004.

[11] A. Pentland, R. Fletcher, A. Hasson, “DakNet: Rethinking Connectivity in Developing Nations”, IEEE Computer, 37(1), 2004.

[12] G. Rossi, S. Murugesan, N. Godbole, “IT in Emerging Markets”, IT Professional 14(4), 2012.

[13] H. Sahilu, A. Villafiorita, K. Weldemariam, M. Belachew, A. Zewge, “Designing Distributed Agricultural Information Services for Developing Countries”, ACM DEV, 2012.

[14] A. Seth, D. Kroeker, M. A. Zaharia, S. Guo, S. Keshav, “Low-Cost Communication for Rural Internet Kiosks Using Mechanical Backhaul”, ACM MOBICOM, 2006.

[15] H. Shirani-Mehr, F. B. Kashani, C. Shahabi, “Efficient Reachability Query Evaluation in Large Spatiotemporal Contact Datasets”, PVLDB, 5(9), 2012.