# All-About Digital Diaries: Opportunities and Challenges

**Laura Ferrari, Marco Mamei, and Franco Zambonelli,** *University of Modena and Reggio Emilia*

**All-about diaries—software platforms that record, in a browsable and machine-processable format, the everyday activities of people, communities, and objects—offer a wealth of application opportunities, but their full-fledge implementation will require overcoming several challenges.**

Tracking personal activities and events in a diary or calendar is becoming a digital process and is assuming a more public dimension owing to social networking tools such as Facebook and Twitter. Furthermore, pervasive computing devices such as wireless sensors and smartphones can now automatically record—in a sharable, browsable, and machine-processable format—the everyday activities of people, communities, and objects.[1,2] Soon, expressive digital diaries about everything—called *all-about diaries*—will emerge. Such diaries present opportunities to both companies and consumers for innovative applications and novel markets.

Although some technologies and proposals for implementing all-about diaries already exist, several challenges remain. Here, we review some of the opportunities and challenges to better understand these emerging systems and the technologies and trends contributing to their development.

## The Vision

After work, Marco puts on his GPS-enabled training shoes and goes for a jog. The shoes automatically upload his running performance to his diary, which his smartphone's sport-tracking application fetches, updating his jogging stats in real time.

After his jog, Marco wants to go out for a drink. His phone's people-like-me application aggregates data from the diaries of people with similar interests and recommends The Fox pub. Marco loads The Fox's diary and sees that he knows many of its regular customers.

While Marco looks through his phone's contacts list, his friend-tracking application connects to his diary and reports that he hasn't seen his friend Franco in a while. A graph shows a

**Table 1. Potential all-about-diary services and those who stand to profit from investing in such services.**

| Business contributors | All-about diary services | | |
| --- | --- | --- | --- |
| | **Personal** | **Social** | **Commercial** |
| **Consumer electronics manufacturers** | Produce sensors and gadgets that can detect user activities and use them to automate and enrich diaries. | Produce sensors and gadgets that can detect proximity and social or collective activities and use them to connect users with friends, colleagues, and acquaintances. | Produce sensors that can detect activities in certain places and interactions with objects. Use them to deliver detailed and contextualized information on products and usage. |
| **Telecom and Web providers** | Host diaries and offer tools that analyze data, identify trends, and send event notifications so users can build, store, and access their personal diaries and basic diary services. | Collect and produce detailed views of social activities and interactions so users can construct global social diaries or special-interest group diaries. | Produce and sell aggregated data on user and social-group behaviors. Use to deliver data on commercial and social trends to enable targeted and personalized ads and marketing analysis. |
| **Application producers** | Produce applications that analyze specific data and use to personalize diaries with specific tools and annotation (for example, for visualizing and extracting data). | Produce applications for special interests groups and offer social matchmaking to various classes of social communities. | Produce applications tailored for specific objects, places, or commercial sectors. Use to offer product recommendations or logistical and planning data to specific industrial and commercial actors. |

recent decline in calls between the two, so Marco calls Franco.

At 9 p.m., Marco and Franco are at The Fox when Marco's phone buzzes: the friend-tracking application recognizes a woman two tables away who jogs in the same park he does. As Marco approaches her, the wall-mounted display nearby recognizes their common interest in jogging (based on information from their diaries) and displays an ad for running shoes. After a while, Marco and the woman share their diaries so they can meet up next time they're both jogging.

This scenario synthesizes many envisioned all-about-diary characteristics, presenting a software platform that

- collects data from sensing devices, software services, and user input;
- executes pattern-analysis algorithms to extract high-level information from the acquired data (for example, correlating accelerometer data with jogging);
- compiles an overall diary (log) with the obtained information; and
- presents its data to several smartphone applications that can offer innovative services.

Diaries can be created for different entities (people, objects, places, and communities), and they can also be aggregated toward social, statistical, or commercial purposes.

From a technological viewpoint, the increasing diffusion of pervasive sensing devices will enable all-about diaries. From a user viewpoint, future acceptance is likely—at least by the digital-native generation (young people eagerly share personal information in social networks). From the social and political perspective, the implications of all-about diaries can be disruptive—like George Orwell's Big Brother. Regardless, the road toward all-about diary applications has already been paved, and companies looking for economic opportunities are now pushing to realize the applications.

## Opportunities

Table 1 shows potential services for all-about diaries as well as who should consider contributing to them as an opportunity for economic or business gain.

### Consumer Electronics Manufacturers

Consumer electronics manufacturers that produce sensors, networking tools, and visualization devices (for integration into everyday objects) can create add-on tools for the all-about diary. Even devices of small intrinsic value, when integrated into the diary, can acquire a high extrinsic value and transfer it to the diary.

Like the iPod ecosystem, where add-ons bring notable value to the music player, the all-about diary can create an ecosystem using sensing devices. For example, consider a galvanic skin-response sensor, which offers a simple way to record emotions. The sensor alone might not seem commercially appealing, but embedding it into an everyday object (such as a watch) and integrating it with a diary would let users capture stress levels and associate them with their locations and activities (see www.biomapping.net). The sensor would add expressiveness to the diary and might help it interpret existing data.

### Telecom and Web Providers
Telecoms and Web providers can offer the computational and networking infrastructure necessary to run and provide access to diary platforms. Offering diary services could attract users or reenforce their loyalty.

Moreover, hosting diaries implies having access to large and finer-grained user data. With this data, the providers could build three-party market systems. In other words, a third party (such as an advertiser receiving user profile information) could pay to participate in the new market created through the free exchange of data between the provider and diary users. The providers could further engage "cross-subsidies" policies, where they sell commercial services enabled by the diary to pay for the cost of offering free user services.

### Application Producers
Application producers can be companies developing software or skilled users creating applications for distribution via suitable application stores. They might be contributing to the diary to use the applications they create or possibly to sell them, or they could engage in the three-party markets and cross-subsidies policies just described.

Application producers can use diary information to create services that offer personalization and context-awareness. Moreover, they can create plug-ins to enrich the diary with tools for data analysis and aggregation. This can generate an applications ecosystem with value beyond the mere intrinsic value of each application.

In addition, diary-based systems will have a profound impact on several specific and important markets and scenarios. For example, in

*The Decision Tree* (Rodale Books, 2010), Thomas Goetz discusses how diary-based systems could affect public health. Chronic diseases (such as heart disease, cancer, and strokes) are sometimes referred as "diseases of lifestyle," because they're strongly affected by our everyday activities (for example, our dietary habits). Applications such as HealthVault (www.healthvault.com) that collect information about such habits can provide invaluable data, which we could use to build new public health policies.

Diary-based applications could also revolutionize the work environment by inferring an organization's social network and the roles its employees play.[3] Who talks with whom? Who's the expert on a given topic? In most organizations, these critical questions are answered only through episodic memories. Instead, management could refer to a reliable log of what's happening in the organization.

> **Like the iPod ecosystem, where add-ons bring notable value to the music player, the all-about diary can create an ecosystem using sensing devices.**

### Existing Technologies
Interest in diary-based services dates back to Vannevar Bush's 1945 essay "As We May Think." Pioneering approaches discussed at the ACM Workshop Series on Continuous Archival and Retrieval of Personal Experiences (CARPE 2004–2006), such as work from Jim Gemmell and Gordon Bell,[4] have furthered that interest, leading to an increasing number of applications. Figure 1a shows the architecture of most of today's systems, which typically

- collect data from a few pre-specified sensors, typically embedded in a smartphone (for example, through the GPS), or from explicit user inputs;
- provide some limited security and privacy mechanisms in the form of access control lists (ACLs) for sharing collected data; and
- provide a Web front end to visualize collected data and a simple API to forward data to third-party sites to use the data as blog entries or mashups.
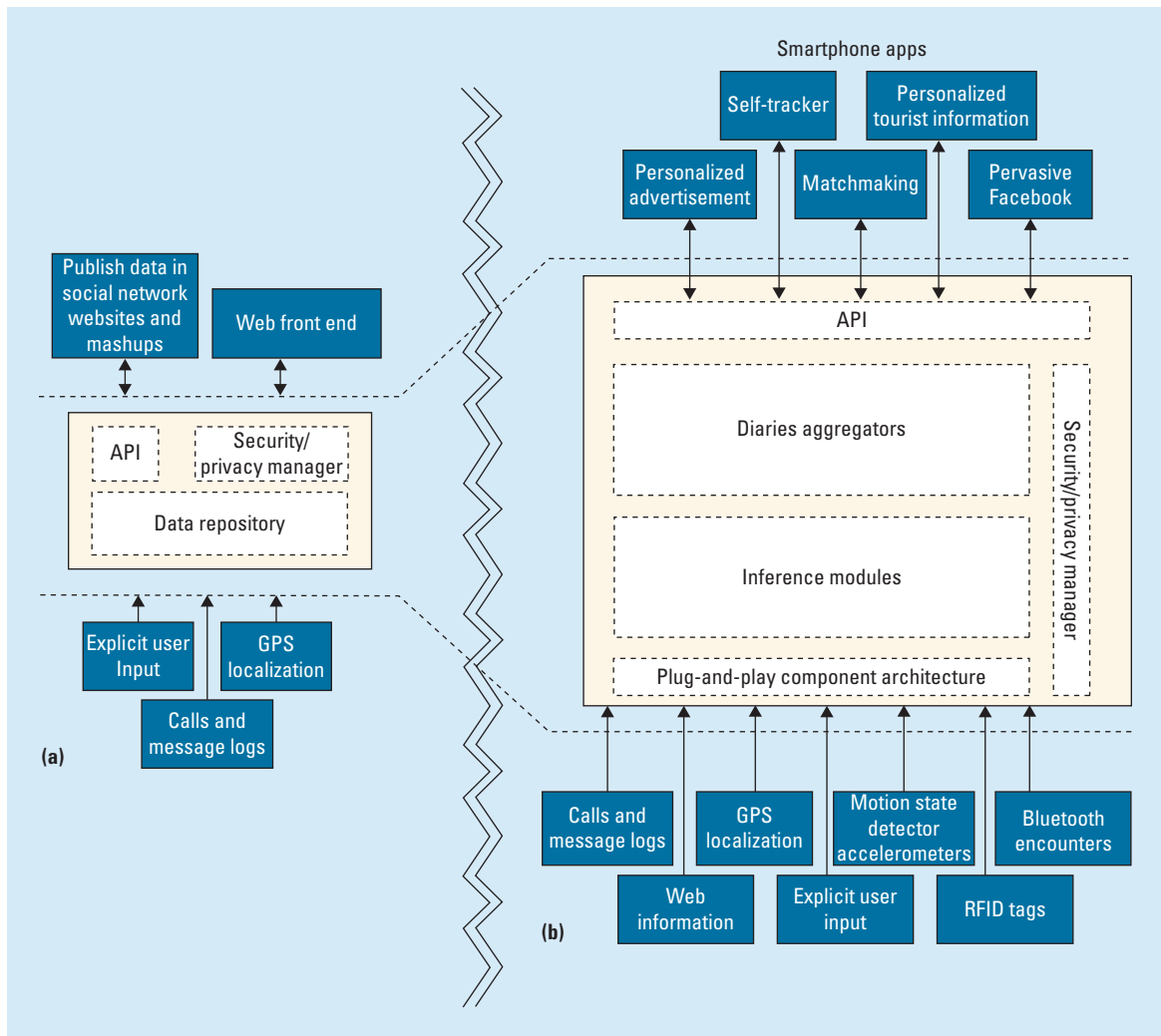
**Figure 1.** All-about diary architectures. (a) The architecture of most available systems; (b) future systems will be open platforms for developing a wide range of applications.

Some such systems are already available to a wide audience.

### Location-Based Diaries
Google Latitude (www.google.com/latitude) is a paradigmatic example of this architecture (see Figure 2a). The application collects location data from user phones and uploads it to a server where information can be shared among selected friends, based on a simple ACL. Foursquare (http://foursquare.com) and Gowalla (http://gowalla.com) similarly share location data.

### Web Self-Trackers
Some websites let users keep a journal in which they enter information about specific activities or issues in their lives. These sites, called Web

self-trackers, exist for monitoring your diet (www.fitday.com), daily habits (http://feltron.com, see Figure 2b), and even sex life (www.bedposted.com). One site (www.quantifiedself.com) collects several proposals in this direction.

These sites can also produce statistics for and summaries of the inserted data. Still, they're not integrated with sensors or portable devices (so they can't automatically update diaries), and they can't extract expressive information from available data.

### Mobile Phone Loggers
The Nokia Life Blog application (www.nokia.com/lifeblog) automatically collects and organizes text messages, notes, and pictures taken from the phone in a diary-like service, and it then posts them on an external blog. Nokia's Vine (www.nseries.com/nseries/nokiavine), which is similar

**Figure 2.** Two representative examples of current diary-oriented systems. (a) The Google Latitude application, which records user whereabouts; (b) a Feltron report of a user's daily habits.

to Google Latitude, could complement this application by adding location information. Similar objectives characterize Context Watcher (http://portals.telin.nl/contextwatcher), which can connect with external sensors.

Although these closely approach our all-about diary vision, they still can't automatically extract expressive information or flexibly and dynamically integrate data sources.

## Future Web Diaries

The current applications we've described were developed in silos—that is, a single Web application logs a few sensors and presents the collected results. We envision future all-about diaries as open platforms for developing a wide range of applications (see Figure 1b).

Several pervasive and environmental sensors connected via wireless and personal-area networks will feed data to the diaries, which will integrate user inputs and Web resources and aggregate and extract patterns. They'll provide the results to external applications via open standards. Several smartphone applications will then access the diary to create specific functionalities, such as those illustrated in our scenario.

The following components will be key to realizing the envisioned diary platform.

### Plug-and-Play Component Architecture

The diary should be a flexible architecture that can host sensors and software modules in a plug-and-play way. If the user buys a device with embedded sensors, the diary should automatically incorporate the new device and its sensors' data.

The diary should also be robust despite unexpected situations, such as device failure or sensor unavailability.

Research in this area includes Microsoft's MyLifeBits, a diary that aims to let users access and index diverse information sources, ranging from wearable sensors to continuous audio and video streams.[1,5] The Betelgeuse project aims to let systems plug in components and sensors dynamically.[6]

However, the recent failure of the Google Wave platform, which attempted to integrate users' Web experiences into a unified architecture, reveals the importance of user perceptions. A coherent and unified platform might clash with user requests for well-defined, simple, and self-contained services.

### Inference Modules

Diaries must extract high-level information from raw sensor readings, such as from

- "location" to "place" (for example, from GPS coordinates to the user's "parents' house");
- "date and time" to "event" (from "25 Dec., 8:20 pm" to "Xmas dinner"); and
- "sensing" to "activity" (from "accelerometer data" to "jogging").

Without these fundamental mechanisms, it's impossible to effectively exploit the acquired data.[7]

Numerous research projects focus on this issue, from proposals for analyzing GPS traces and extracting the places visited[8] to techniques for effectively recognizing user activities from body-worn accelerometers.[9] Researchers are

also exploring activity recognition systems that receive data from embedded cameras (for example, SenseCam[10]) and data-mining techniques to infer a person's activities from his or her interactions with RFID-tagged objects.[11] There are also data-mining techniques that, based on proximity sensors (such as Bluetooth) or phones logs, can infer social relationships between people.[12]

### Diary Aggregators

Researchers have developed several techniques, including collaborative filtering and recommendation systems, for aggregating information online from multiple users. A few projects (www.citysense.com) are exploring data aggregation in the context of diary-like applications, attempting to extract relevant features by collecting mobility patterns.[13] Extracting aggregated information from a collection of diaries could lead to several new services. For example, a restaurant might be able to infer its average clientele by aggregating its customers' diaries.

> Cultural change and evolution will strongly influence the diffusion of diary-based applications.

### Challenges

Realizing the concepts just outlined and establishing widespread acceptance present many challenges.

### Security and Privacy Concerns

Security and privacy are of primary importance for diary-like systems. Security implies preventing unregulated access to diaries and proliferation of mock-up data that biases collective analysis. Assessed techniques (ACLs and Captcha-based solutions) can help address these issues. Privacy, on the other hand, implies that users (and secondary stakeholders, such as people observed in someone else's diary) can control what information they share and with whom. This presents more difficult challenges.

Current solutions (mainly relying on simple data anonymization) don't fully account for the complex issues that can arise as inference and aggregation operations are applied. For example,

it might be acceptable to share blurred GPS coordinates ("Marco was near latX, lonY") but not the inferred location (Marco was at "The Fox" pub). Similarly, there might be situations in which raw data shouldn't be shared (from a biosensor capable of identifying health problems) but knowledge inferred from it should be (for example, if the biosensor only infers an emotion like happiness). Consider the DARPA LifeLog project, which was cancelled in 2004 because of privacy concerns.

### Open Standards

This isn't a niche issue for diary systems. The need for open architectures and standards is a general issue of modern networked scenarios, as are the associated challenges of

- defining open means to release self-describing components,
- identifying general-purpose data and metadata models that can capture the main characteristics of sensors and data sources, and
- finding the proper trade-off between the need to adopt common ontologies for data and metadata and the necessity to pragmatically account for emerging ontologies.

An interesting proposal in the context of sensor data integration advocates using industry standards such as OSGi (http://osgi.org) and CIShell (http://cishell.org) to construct flexible architectures.[14] These standards, however, deal only with the syntactic aspects of data and service integration. Semantic integration represents an open research challenge.

### Cultural Evolution

Cultural change and evolution will strongly influence the diffusion of diary-based applications. From a social perspective, the massive adoption of social networks is the beginning of a cultural change that will foster the adoption of diary-based systems. Today, it's almost rude to be unavailable by mobile phone for a long time. In the future, it might be impolite to not share diary information with your social network.

From a personal perspective, people will start mining their diary information for entertainment, self-reflection, and guidance. The Feltron annual

report (recently acquired by Facebook) is a paradigmatic example of this idea (see Figure 2b). Similarly, several applications are already available to visualize graphs and statistics about a user's friendship network, call logs, and SMS logs. As these tools gain widespread adoption, all-about diaries will start to emerge.

All-about diaries promise to be key components of our future networked society. Despite the many challenges we face in realizing such systems, they represent a gold mine for novel applications, presenting numerous business opportunities. ⒤⒯

## Acknowledgments

## References

1. G. Bell and J. Gemmell, *Total Recall: How the E-Memory Revolution Will Change Everything*, Dutton, 2009.
2. T. Maekawa et al., "Object-Blog System for Environment-Generated Content," *IEEE Pervasive Computing*, vol. 7, no. 4, 2008, pp. 20–27.
3. D. Olguin and A. Pentland, "Sensor-Based Organizational Design and Engineering," *Int'l J. Organizational Design and Eng.*, vol. 1, no. 2010, pp. 69–97.
4. J. Gemmell et al., "Passive Capture and Ensuing Issues for a Personal Lifetime Store," *ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, ACM Press, 2004, pp. 48–55.
5. J. Gemmell, G. Bell, and R. Lueder, "MyLifeBits: A Personal Database for Everything," *Comm. ACM*, vol. 49, no. 1, 2006, pp. 88–95.
6. J. Kukkonen et al., "BeTelGeuse: A Platform for Gathering and Processing Situational Data," *IEEE Pervasive Computing*, vol. 8, no. 2, 2009, pp. 49–56.
7. A. Sellen and S. Whittaker, "Beyond Total Capture: A Constructive Critique of Lifelogging," *Comm. ACM*, vol. 53, no. 5, 2010, pp. 70–77.
8. L. Ferrari and M. Mamei, "Discovering Daily Routines from Google Latitude with Topic Models," *Proc. IEEE Workshop on Context Modeling and Reasoning*, IEEE CS Press, 2011, pp. 432–437.
9. N. Bicocchi, M. Mamei, and F. Zambonelli, "Detecting Human Activities from Body Worn Accelerometers via Instance-Based Algorithms," *Elsevier Pervasive and Mobile Computing*, vol. 6, no. 4, 2010, pp. 482–495.
10. D. Nguyen et al., "Encountering SenseCam: Personal Recording Technologies in Everyday Life," *Proc. Int'l Conf. Ubiuqitous Computing*, ACM Press, 2009, pp. 165–174.
11. M. Buettner et al., "Recognizing Daily Activities with RFID-Based Sensors," *Proc. Int'l Conf. Ubiquitous Computing*, ACM Press, 2009, pp. 51–60.
12. N. Eagle, A. Pentland, and D. Lazer, "Inferring Friendship Network Structure by Using Mobile Phone Data," *Proc. National Academy of Sciences*, PNAS 0900282106, vol. 106, no. 36, 2009, pp. 15274–15278.
13. Y. Zheng et al., "Recommending Friends and Locations based on Individual Location History," *ACM Trans. Web*, vol. 5, no. 1, 2011, pp. 5:1–5:44.
14. K. Börner, "Plug-and-Play Macroscopes," *Comm. ACM*, vol. 54, no. 3, 2011, pp. 60–69.

**Laura Ferrari** *is a PhD student at the University of Modena and Reggio Emilia. Her research interests include pervasive computing, location-based services, and data mining. Ferrari received her MS in management engineering from the University of Modena and Reggio Emilia. She's a member of IEEE. Contact her at laura.ferrari@unimore.it.*

**Marco Mamei** *is an assistant professor at the University of Modena and Reggio Emilia. His research interests include pervasive computing and context awareness. Mamei received his PhD in computer science from the University of Modena and Reggio Emilia. He's a member of IEEE. Contact him at marco.mamei@unimore.it.*

**Franco Zambonelli** *is a full professor at the University of Modena and Reggio Emilia. His research interests include autonomic and pervasive computing and complex and adaptive systems. Zambonelli received his PhD in computer science from the University of Bologna. He's a senior member of IEEE and the ACM. Contact him at franco.zambonelli@unimore.it.*

**cn** **Selected CS articles and columns are available for free at http://ComputingNow.computer.org.**