Identity in the Information Society

10.1007/s12394-010-0047-x

SmartData: Make the data “think” for itself Data protection for the 21st century

George J. Tomko[1], Donald S. Borrett[2], Hon C. Kwan[3] and Greg Steffan[4]

(1) Identity, Privacy and Security Institute, University of Toronto, Toronto, ON, Canada
(2) Division of Neurology, Toronto East General Hospital, Toronto, ON, Canada
(3) Department of Physiology, University of Toronto, Toronto, ON, Canada
(4) Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, Canada

George J. Tomko
Email: gjtomko@hotmail.com

Received: 23 December 2009 **Accepted:** 2 February 2010 **Published online:** 27 April 2010

Abstract
SmartData is a research program to develop web-based intelligent agents that will perform two tasks: securely store an individual’s personal and/or proprietary data, and protect the privacy and security of the data by only disclosing it in accordance with instructions authorized by the data subject. The vision consists of a web-based SmartData agent that would serve as an individual’s proxy in cyberspace to protect their personal or proprietary data. The SmartData agent (which ‘houses’ the data and its permitted uses) would be transmitted to, or stored in a database, not the personal data itself. In effect, there would be no personal or proprietary “raw” data out in the open—it would instead be housed within a SmartData agent, much like we humans carry information in our “heads;” extending the analogy, it would be the “human-like clone” that would be transmitted or stored, not the raw data. The binary string representative of a SmartData agent would be located in local or central databases. Organizations requiring access to any of the data resident within the agent would query it once it had been “activated.” In this paper, we provide a preliminary overview of the SmartData concept, and describe the associated research and development that must be conducted in order to actualize this vision.

Keywords SmartData - Intelligent agents - Security - Privacy - Data protection

This paper is published under the auspices of the Identity, Privacy and Security Institute at University of Toronto, Canada

---

Introduction

SmartData is a research program to develop web-based intelligent agents that both house an individual’s personal and/or proprietary information, and protect the privacy and security of the data. It is presently the subject of a multi-disciplinary research program in the Identity, Privacy and Security Institute (“IPSI”) at the University of Toronto, Canada. The vision is that each person or designated entity would

possess its own surrogate web-based SmartData agent as the entity which houses sensitive information and which would be transmitted to, or stored in a database—not their personal or proprietary information. A SmartData agent would replace the need for the storage and transmission of “raw, sensitive data” and, moreover, would serve as an individual’s proxy in cyberspace to protect the privacy and security of their personal or proprietary information. In effect, the personal or proprietary data would be empowered to protect itself from secondary or unauthorized uses.

But how, one might ask, can data protect itself? The short answer is, by transforming personal data into an “active” form—namely transforming it into “SmartData.” If stated as a question, the analogy would be: What if data were part and parcel of an intelligent agent which, just like a smart human, would only divulge its personal information when it was safe to do so? What if its decisions were based, not only on the instructions set out by its surrogate owner, but additionally, by the content and the context of the situation, that is, the need for the personal information, relative to the background situation? We posit that the ability to make decisions in consideration of the background situation or broader context is mandatory; otherwise web-based agents will be “brittle” in nature and have limited usefulness.

Although SmartData’s initial focus will be to protect information within the current “flat” web-based environment comprised largely of text and images, the design methodology will also accommodate what we view as the next technological evolution in information sharing—namely, sharing information in three-dimensional virtual worlds as currently pioneered by companies such as Linden Research in Second Life. Just as the original Internet was one-dimensional and only processed text, the introduction of the World Wide Web brought about a second dimension allowing both text and images to be shared. As bandwidth improved, this enabled “moving images” or video to be transmitted as well. With technology advancing dramatically, Philip Rosedale, inventor of Second Life, makes a compelling case that the current flat internet will be transformed into a 3-D virtual world wherein text and images will only form a subset of the total environment. He argues that since humans evolved in a 3-D world and are by nature social animals, a corresponding virtual world would allow more familiar, efficient and social ways of exchanging information between individuals’ avatars, but at the “speed of light.”

There is, however, another aspect to consider. To date, the user has always been “external’ to the Web in that he/she interfaces with the Web through the use of a keyboard, in its simplest form, or via a surrogate such as a computer programmed to carry out the user’s instructions. The user, however, remains on the “outside looking in.” A similar situation exists in current 3-D virtual worlds such as Second Life wherein “one’s avatar” is, for the most part, directed by a human or computer-surrogate on the outside, in the “real world.” Getting “inside” the Web has now begun, although mainly for malicious purposes, with the introduction of “dumb” agents such as viruses, worms, cookies, and Trojan horses. Here, these agents are essentially dumb in that once inside the Web, they can only take actions based on previously programmed code or algorithm—they have no agency, per se. What is important to note, however, is that this development was the first to establish the evolutionary direction of “acting-agents” moving “inside” the virtual world.

We propose that the next transformation will be the introduction of intelligent, embodied agents within a 3-D virtual world, wherein those agents will become our acting-surrogates. Such a “virtual web-world” could spawn a far more secure, efficient and productive way of exchanging and processing information, and inspire totally new innovations, as did the flat web. However, the demand for privacy and security in such an environment will escalate dramatically. In the spirit of Privacy by Design[1], this research program will undertake to “build-in privacy and security” as a single constituent, right from the outset. The direction of this program will be to develop SmartData virtual agents to protect both privacy and security, in conjunction with the development of the next generation 3-D virtual “web-world.” In other words, SmartData would be poised to form an integral part of the infrastructure for the entire 3-D virtual web-world. It is important to start this process now in the initial planning and design stages of the

research and development in order to ensure that the feature of privacy protection is included and give equal priority to security.

Part of the strategy of our research program is to work towards having the virtual world environment positioned as a hub on a new experimental high-speed Internet II developed by the GENI project (Global Environment for Networked Innovations). This high bandwidth platform would more easily allow for the development of SmartData to be carried out using the concept of open or “crowd-sourcing” wherein its development would be opened up to interested researchers around the world, who would be encouraged to try out their own ideas, in combination with other investigators. By establishing a global interdisciplinary cross-fertilization process, we believe that there will be a far greater chance of developing truly cognitive agents than through a few researchers, working on their own, in isolated labs.

The direction of our research is based on the premise that natural evolution remains the best roadmap available for building artificial agents possessing cognition. Specifically, we are proposing a hybrid methodology whereby we will structure and program populations of diverse simple agents, each initially with a different top-down design, wherein the structural or algorithmic variables of the designs are reflected in the agents’ chromosomes. These populations of agents will then be placed in a virtual evolutionary environment which includes competition and cooperation with other agents such that their selection leads to higher levels of fitness, as measured by our specification for SmartData—the fitness criteria for survival of evolving agents. We acknowledge that the “cognition” property achievable in our initial attempts may be trivial in nature, compared to that of a human. However, we believe that by developing the right evolutionary line of attack to deliver even “trivial cognition” will represent a major step forward in evolving more complex forms of cognition, suitable for protecting the privacy and security of data.

Our view is that the evolution of cognition in agents is as much the task of choosing the right dynamic world, as choosing the right “neural control system.” It is the world and its particular complexity which appears to determine which set of emergent properties in an agent will enhance survival, and are therefore selected. Function, in this case, dictates structural design. This principle, we maintain, works all the way up to cognition. Accordingly, a major component of this research will be to create the right dynamic virtual environment within which to evolve these agents. The stipulation of a complex virtual world will eventually require a high speed simulation computer for the virtual world and high speed network connections between the physical embodiment of the agents and the virtual world simulation computer in which they are graphically represented.

SmartData Concept

1. Storm Clouds on the Horizon
   The introduction of cloud computing has highlighted the difficulty of securing privacy and protecting personal information in a Web 2.0, world especially when there are competing forces interested in accessing the personal information. Governments and public officials want unrestricted access to information in the name of public safety, crime prevention, and formulating future policies and legislation. Businesses would like to have unfettered access for the purposes of marketing, advertising, improving customer service, and at times, to use as a revenue generator, by selling lists of personal data to other organizations. Consumers and citizens generally want to divulge their personal information for specific purposes, after which the data should be destroyed since many are increasingly becoming aware of the threat of

identity theft, if their data ultimately fall into the wrong hands. But the difficulty with current data protection schemes, caught in the tug-of-war of competing interests, is that once the data is in plain digital text, it can be easily copied and distributed against the expressed wishes of the data subject. Even if the data are encrypted and married to digital rights management protocols, it must eventually be decrypted for actual use.[2] Personal information, once released for a singular purpose, may now become lost forever in the “cloud,” potentially subject to unending secondary uses. The difficulties are highlighted in the following excerpt from the EU Tender Specification, *“The cloud: understanding the security, privacy and trust challenges”*[3]:

> A commonly known concern stems from the lack of clarity (or diverging views) on who is responsible for the data residing or moving in the cloud and under which jurisdiction it falls. One possible method is that cloud data are governed according to the legal framework of the country of origin through establishment of data-controllers who are responsible for compliant handling of data under their control. Such controllers could transfer control to someone else under another jurisdiction under certain rules. Another approach could be to choose the jurisdiction of the country of storage. The former introduces the issue of the policy or terms of service that cloud providers offer to users. For instance, who owns the data, who can use them and who is responsible for data processing and giving access to the data, and under which conditions?
>
> Cloud providers may have themselves access to sensitive data stored and processed in their clouds; also, they might reserve the right to disclose stored data to third parties, or reserve the right to modify modalities at a later stage, depending on the law they function under. Even if a cloud provider assures protection of data stored in his own cloud, he might be using another cloud provider or at a certain moment be taken over by a cloud provider with different assurances, rendering possibly the data accessible to third parties including competitors. This makes contractual provisions and assurances on data controller responsibilities critical. Wrong arrangements may lead to situations that constitute data protection violations in the users’ jurisdiction, but might de-facto render any legal or accountability claim void.
>
> The above mentioned concerns are mainly stemming from the implementation of legislation and policies. However concerns may also stem from the technology and infrastructure that is used to implement the cloud services, such as encryption, virtualization, identity management and access control. The mentioned concerns are often insufficiently known by users or if they are known, they may hinder the development of cloud services due to lack of trust.

These difficulties, although tempered by regulatory policies and legislation, can never be completely surmounted because their source arises from the way in which data has been treated since the advent of digital databases. Data has essentially been treated as passive in nature and, in effect, this is true. But not only is it passive, it is also “dumb” in the sense that it simply sits on a storage device or is sent from point A to point B. At its root, this is the precise problem we are facing: the personal or proprietary information of an entity, be it medical or financial in nature, or a trade secret—as represented by a binary string of data residing in the cloud, is not capable of protecting itself against unauthorized, secondary uses.

In an attempt to overcome likely infringements, a tangled web of international legal provisions and commercial contracts has been established to address the various privacy and proprietary concerns of different jurisdictions. A potential consequence is that, in attempting to protect personal information, a greater portion of what to-date, has effectively been located in the intellectual public domain will hence be enclosed in private domains, with the downside of restricting future innovation and creativity. Furthermore, the move to legislate personal information as a property right is also fraught with jurisdictional difficulties and potential political battles over its appropriateness. The prospect of what we may face is not bright—not only a global legal bureaucratic nightmare but also a technical morass of different systems and standards, all trying to interface with each other. While no system can solve all of these problems, we propose that the goal of making data self-sufficient, in effect, capable of protecting itself, will circumvent many of the legal and technological issues. One benefit we see is that the regulatory framework and legal structures between parties need no longer to be the first-line of defense: they will be transformed into serving as the backstop, in the same way that many retail establishments use laws against theft of merchandise as a secondary line of defense—with the primary line of defense being a secure building, the installation of anti-theft systems, the presence of security staff, guard dogs, etc.

Thus far, a “zero-sum” approach has prevailed over the relationship between technology and privacy. A zero-sum paradigm describes a situation in which one party’s gains are balanced by another party’s losses—win/lose. In a zero-sum paradigm, enhancing usability and functionality necessarily comes at the expense of privacy; conversely, adding user privacy controls would be viewed as detracting from system performance. Rather than adopting a zero-sum approach, we propose a “positive-sum” paradigm, which we believe is both desirable and achievable, whereby adding privacy measures to systems need not weaken security or functionality but quite the opposite, would serve to enhance the overall design. A positive-sum paradigm describes a situation in which participants may all gain or lose together, depending on the choices made—win/win, not win/lose.

2. Moving from “dumb” to SmartData

SmartData is a binary string representing the neural architecture of the virtual agent. The string will contain both the code for what we will call “intelligence” and also the personal or sensitive information incorporated within the neural architecture. As such, there will be no different pieces of the binary string; it will not contain an identifiable segment representing the sensitive data. The *entire* string will specify the neural and somatic (if needed) architecture of the agent. This binary string or code, when downloaded into a hardware device that is reconfigurable, such as a Field Programmable Gate Array (“FPGA”), will restructure the device into the architecture specified by the binary code, and “activate” a unique agent. The analogy to biological agents, such as humans, is that the personal or sensitive information is stored within the memory of the agent (i.e. contained within the connections of its neuronal network). By incorporating the sensitive data into the neural architecture, the agent or SmartData can serve as a proxy of its owner to either protect or release its data, based on its owner’s instructions and the background situation. In other words, the data becomes “smart” by virtue of its being inextricably incorporated and secured within the intelligent agent—in effect, it takes the form of memory in an eidetic agent.

Each agent, as a surrogate of an individual or some specific entity, will be unique in its neural connections, just as the neural connections in human brains are unique. As part of the research underlying SmartData, we will examine how this is best configured—as a single, centrally-located proxy, or as multiple, domain-specific agents. In the case of the latter, the copies will in effect become clones. The question we must answer with clones, given that they will be in

separate domains and undergoing continual learning (a necessity if the context of a situation is to be a factor in decision making), is whether the clones are allowed to get out of “sync” or whether we have to occasionally allow them to “reconnect” and re-sync. This may be comparable to identical twins getting together at the end of a day and exchanging what has happened, and what is new with them. Re-syncing has benefits in terms of maintenance. If one of the binary strings for a particular clone becomes corrupted, then another clone can take its place. Although some of these issues may appear academic and “distant” at this point, we maintain that their consideration is important in the initial planning and design stages to ensure that privacy protection is included as an essential feature and not treated as an afterthought, yet again.

Personal information, under our scenario, would no longer be stored as an exposed string of binary digits, encrypted or otherwise, on a storage device somewhere out there, possibly in the cloud. Instead, the SmartData binary string for each person would be stored in the cloud. This concept is sympathetic to the growing recognition among information and security professionals that traditional ways of securing data are inadequate—whether they are stored in proprietary systems or “in the cloud.” Building bigger and better walls around ever-larger databases is simply not feasible. It is becoming widely recognized that safeguards must be applied early and “in depth.” The logical end-result of this trend is to shrink the security and privacy perimeters down to the data item-level, as advocated by the Jericho Forum and Cloud Computing Security Alliance[4]. We propose to go even further—to make the data itself, intelligent and active.

The resulting feature is that “raw” personal data would never be transmitted—only the binary string representing the unique embodiment of the agent would be disclosed. As an example, consider a personal electronic health record. The health information would be stored within the binary string representing an individual’s SmartData and placed on a storage device in the cloud, along with other individual SmartData strings. When a request to view a health record is received, its SmartData binary string would be downloaded from the cloud into some type of reconfigurable hardware (whether that hardware needs to be a secure trusted-hardware device such as that incorporated into the Trusted Platform Modules has yet to be determined), which would be reconfigured based on the instructions in that specific binary string. This would “activate” the SmartData agent and allow it to be queried for specific medical information. Should the request for medical information be valid, SmartData would then release the information in accordance with the relevant contextual factors. These factors would include, for example, intended purposes, identity, authentication and authorization (including strength of reputation and/or trust); the policies and practices in place; and any other conditions, legal or otherwise. At this point, depending on its implementation, the data subject might be made aware of the permitted access; this notification could also take other forms, such as periodic (perhaps monthly) audits, regular Twitter reports or its equivalent in the future.

Another option that may be suitable in some circumstances is only to expose personal or proprietary data in analog form (such as music), or image format (for LCDs). In these circumstances, the digital-to-image or digital-to-analog format conversions would be performed within the code of the SmartData agent, which would itself be embedded in a secured trusted-hardware device such as an FPGA. We envision, at this early stage, that these reconfigurable hardware devices will be USB-like compatible allowing for attachment to any device such as a computer or PDA, where there was a need to access sensitive information. The secure hardware devices could also be made resident in display devices such that personal information first be decrypted within the secured device, and only the “image format” text actually released[5]. However, the overriding precaution must be that even information in plain-text analog or image format would only be released on a “need-to-know” basis with proper

authentication. In other words, the entity to which the information is released must be trusted and authenticated.

In situations where all the personal information about patients needed to be transferred, such as a patient transferring from one hospital to another hospital, it would be the database of individual SmartData agents (i.e. clones) that would be transferred. Accessing personal information on a particular patient would require the relevant SmartData binary string to be downloaded into a secure reconfigurable hardware device to activate the agent. Again, after the requesting party and the SmartData agent performed mutual authentication, the specific information requested would be presented in analog or image format. Accordingly, personal information would never be exposed until a specific request was made, and then only in analog format. In certain cases, for example, where multi-patient health information in digital format is required for research purposes or outcomes management studies, the specific request would be made to the SmartData agents representing the group of patients in question, and after successful authentication, each SmartData agent would provide a de-identified copy of the information in unencrypted digital format[6].

3. The Security Model

The sensitive data residing “within” SmartData will be structured into two lines of defense. The data will first be divided into logical segments or “lock-boxes,” with each lock-box encrypted using standard cryptographic techniques where the encryption key(s) are analogous to the locks. Each lock-box will have a “non-personal” description of its data as an attached meta-tag; for example, as clinical laboratory results, medical imaging data, etc. With the contents of the lock-boxes encrypted, the SmartData agent would only “know” the meta-tags and the fact that there was a lock-box of encrypted (untranslatable) information associated with each of them. Such methods of traditional encryption will be the first line of defense. However, we will also experiment with a second line of defense based on the properties of nonlinear dynamical systems as exhibited by neural networks. Here we provide a brief overview of the approach. Neural networks exhibit the property that information is not stored in one specific location but across the entire network, manifesting itself as an aggregate of small changes in all the neurons of the network. Similarly, the meta-tags and encrypted information would be stored across SmartData’s neural network. Critical, however, from a security perspective, is that the lock-boxes would be “mixed in” with the agent’s other information or “knowledge”, unrelated to the encrypted data. This would result from the agent learning, or more correctly, memorizing the meta-tags and encrypted contents of the lock-boxes.

For many reasons, including the necessity to evolve a type of “language” or communication code, each SmartData agent would be evolved within a community of agents, and by necessity, each would have a different perspective from that of any other[7]. As a result, these differing perspectives would give rise to different sensory inputs which would modify the underlying neural substrate within which the information is stored. Therefore, the same information stored by two different agents would be represented by a different aggregate of changes in their corresponding neural networks. This information would be “situated” within a hierarchy of nonlinear dynamical systems as represented by the neural architecture, and the details of that architecture will be unique to each agent as a result of their differing perspectives both in the present and throughout their temporal history. Accordingly, the method of storage used, even for identical personal information, would be unique to each agent.

The above has important security implications. As noted, every agent would be unique since no two SmartData agents would store the same data in the same way. Not only will each agent store its data differently, but each interaction will modify the structure of its neural network, which will in turn, change the way that the data is stored. In other words, the result of each

interaction or experience will “shuffle” the locations of the meta-tags and associated lock-boxes. This “shuffling” of personal data as a result of experience, combined with the encryption of the personal data by traditional cryptographic methods, will, we predict, provide double immunity from outside attacks. The binary string representing the SmartData agent (which stores the personal data of an individual) would change over time, with each interaction; and how it would change would be a function of the unique history or experiences of each agent (in essence, its “life”). It would be analogous to a single-use password system that keeps changing with the experience of the user, such that the “key” to decrypt the storage pattern of each agent would in effect become its “life history.”

A personal story which may serve as an analogy in helping to understand this process from SmartData’s perspective is the following: Many years ago as an alter boy in the Catholic Church, I memorized the Latin prayers required to serve mass. I did not then, nor do now, understand Latin. However, to this day, many years later, I can still recite the prayers verbatim—in Latin. Having had many different experiences throughout my life, those prayers have been “shuffled” around the neuronal network of my brain (as a function of my life history), but I can still access them because of their meta-tag, “Latin Prayers learned as Alter Boy,”, which causes me to immediately recall that information. Now, if someone were to provide me with a Latin-to-English dictionary, and rules for translation, akin to Searle’s Chinese Room Argument, I could translate the prayers into English. The Latin-to-English dictionary could be regarded as the decryption key. Depending on the length and complexity of the prayers, once the dictionary (key) was no longer available, I would again have little idea what the prayers meant in English.

To summarize, the encrypted segments and associated plaintext meta-tag will be embedded into the structure of a dynamical system as represented by the neural network within the Reconfigurable Computers. Each agent will store this information in a differing fashion, as a result of possessing a different perspective from that of any other agent. The structure of the neural network will change in a unique manner, as each agent evolves and mutates based on its own experience of the virtual world. Accordingly, we believe there will be no single means of attack applicable to all agents. We propose that this property of possessing differing perspectives, in conjunction with a neural dynamical system, will provide greater security than traditional methods, and we will seek to confirm that hypothesis.

4. Potential Security Attacks on SmartData

If the binary string associated with SmartData was downloaded into an attacker’s PC and analyzed offline, the task of finding the location of the meta-tags and encrypted lock-boxes would be analogous to separating the wheat from a mountain of chaff where both the wheat and the chaff were identical in appearance! SmartData’s encrypted binary strings, within the mountain of unencrypted binary data, would not be statistically different in appearance. Moreover, there would not be a “location” for the personal information since the information contained in a neural network is a function of its structure and embedded in the n-dimensional phase space trajectories (where n is the number of neurons). As mentioned, information in large neural networks is stored within the set of nonlinear dynamical systems. The trajectories of the various dimensional dynamical systems serve as inputs to one another, in a circular feedback. An output is only produced if SmartData is “awake,” (so to speak), after proper authentication has taken place (in effect, a proper set of context-dependent initial conditions), and proper decryption key(s) have been furnished. Although the physical structure of the neural control system embodied in the reconfigurable computer portion of the device could be garnered from the SmartData binary string, that knowledge would not in itself, provide any information about the location or identity of any type of data, encrypted or otherwise. However, if the SmartData

was “inactive,” the structure of the neural network alone would be inadequate to determine the information contained within. Since the information is only embedded in the dynamical systems, they can only be enacted if the SmartData is “active.”

Performing “surgery” attacks by specifically altering single connections or bits in sequence will also not generate any useful data. The nonlinear dynamical neuronal systems comprising SmartData must have the property of structural stability and redundancy to ensure that reliable behaviour and retrieval of correct data continues, even though there are minor perturbations to the system. But the properties of dynamical systems are also such that at the limits or thresholds of structural stability, the system becomes sensitive to small differences in input or structure. At these limits, small differences in the neuronal inputs or connections can lead to large differences in the way that nonlinear dynamical systems behave or store information. Therefore, initially there will be little if any change to surgical sequence attacks because of structural stability, followed by unpredictable large changes beyond the thresholds.

For the reasons discussed above, similarities in structure across different SmartData agents could not be used as a formal attack either. A security feature of SmartData is its property of uniqueness. The “location” of the encrypted binary strings within the dynamical systems would be unique for each SmartData agent because of the different life history and perspectives it undergone during its evolution and learning phase. Indeed, the location of all the stored information and the concomitant structure of the neural control system (comprised of the number of neurons and connections), would be unique.

Furthermore, in traditional cryptography, the underlying premise which allows a hacker to break the encryption of a key or cipher-text is that the encryption algorithm is known. Without the algorithm, the encryption of a reasonably complex cipher could never be broken. Currently, a security system based on keeping the algorithm secret is considered tenuous since the assumption is that, since it is knowable, eventually it will be exposed and then the potential of untested flaws in the algorithm will be exploited. Publishing the algorithm allows the cryptographic community to try different methods to break the encryption, and to the extent that they are unsuccessful, the robustness of the encryption method in question is supported. With SmartData however, the shuffling “algorithm” is, for all practical purposes, unknowable! This fact further enhances the potential security of SmartData.[8]

In summary, SmartData would serve to transform the existing zero-sum mentality, which remains the current-day paradigm underlying the debate around security versus privacy. SmartData is the ultimate example of Privacy by Design whereby the old-world acceptance of zero-sum can now be replaced by the next wave of positive-sum. Most important, the individual would regain control over the uses of his or her personal information, while businesses could still use their customers’ information, but only for permitted purposes. Public officials could also continue to gain access to critical personal information, but only for intended purposes, if and only if it was agreed to by the individual or, in covert situations, sanctioned by the courts. All of this could be done in an efficient manner, serving to remove much of the bureaucracy which currently permeates the process. This is the vision we are advancing in the development of SmartData.

Research Strategy

Clearly we do not possess at this time, nor will we in the near future, enough specific knowledge to design an appropriate algorithm such that artificial agents exhibit expertise in an area as broad and

dynamic as privacy and security. This doesn’t mean that expertise or intelligence is not computational; just that all current approaches are problematic. However, if we use humans as an example of the kind of expertise we wish to achieve, then the exemplar of a proven methodology is natural evolution. Beginning with the simplest of organisms, all life forms and their associated expertise and intelligence, from insects to humans, have all evolved from the “bottom up.” In fact, humans are the existence proof that evolution works—and works very well. The problem is that even though evolution is an amazing problem-solver and novelty engine, it would take far too long to evolve expert agents from the bottom up in the “natural way.” It took roughly three billion years to evolve from a single cell to a human being, and obviously we don’t have the luxury of time. It is critical that we have the ability to rapidly evolve a diversity of agents simultaneously in the virtual world, so that a large number of generations can be processed in a reasonably short period of time. Evolving “physical” agents in the “real world” and in “real time” is clearly not feasible—life is simply too short!

By duplicating the process of evolution within an appropriate environment, using a suitable substrate as our “proto-agents,” and by speeding up the evolutionary process to take advantage of the novelty generated by the entire expanse of the evolutionary tree (without the necessity of pruning for reasons of time constraints), we believe we ultimately will be able to evolve agents which possess cognition. It is only now through the tremendous advances in simulated virtual worlds, developed by the digital gaming industry and platforms such as Second Life, that evolution in a three dimensional virtual world is now possible. Therefore, a critical step in the evolution of intelligent agents is the construction of a proper apparatus which can serve as a test bed or “experimental sandbox’ where different models may be tested. Our long-term objective is to design a system that contains a simulated virtual world, populated by complex virtual agents which compete and rapidly co-evolve through large numbers of generations. SmartData is divided into five research projects as detailed in the accompanying Appendix.

Conclusion

In the future, an additional arsenal of tools will be needed to protect privacy and security in the online world. With the advent of Web 2.0, 3.0 and the Semantic Web, with more and more activities taking place in networked platforms in the Cloud, with remote computing services further removing the common touch points we are presently familiar with, current-day protections will become largely ineffective. Our existing reliance on regulatory compliance and after-the-fact methods of seeking redress will no longer be sustainable, for one simple reason—they will no longer serve as effective means of data protection. Therefore we must look to the future and create futuristic means of protection. This is the entry point for our vision of SmartData—empowering virtual, cognitive agents to act on our behalf to protect the data entrusted to them. In order to achieve this goal, many disciplines will be called upon and the efforts of researchers from multiple fields will need to be tapped. While the goal may at times appear to be daunting, we believe it will be well worth the effort. If we can teach virtual agents to “think” on our behalf, in limited circumstances, that serve to protect the data entrusted to their care—that would represent a significant step forward in ensuring that privacy could live well into the future. It would be the ultimate embodiment of Privacy by Design, which insists upon taking proactive measures to protect privacy and security, in a positive-sum manner. That is the essence of SmartData—an innovative approach to protecting privacy and security, by building cognitive agents to protect our data.

Appendix

SmartData Research Plan

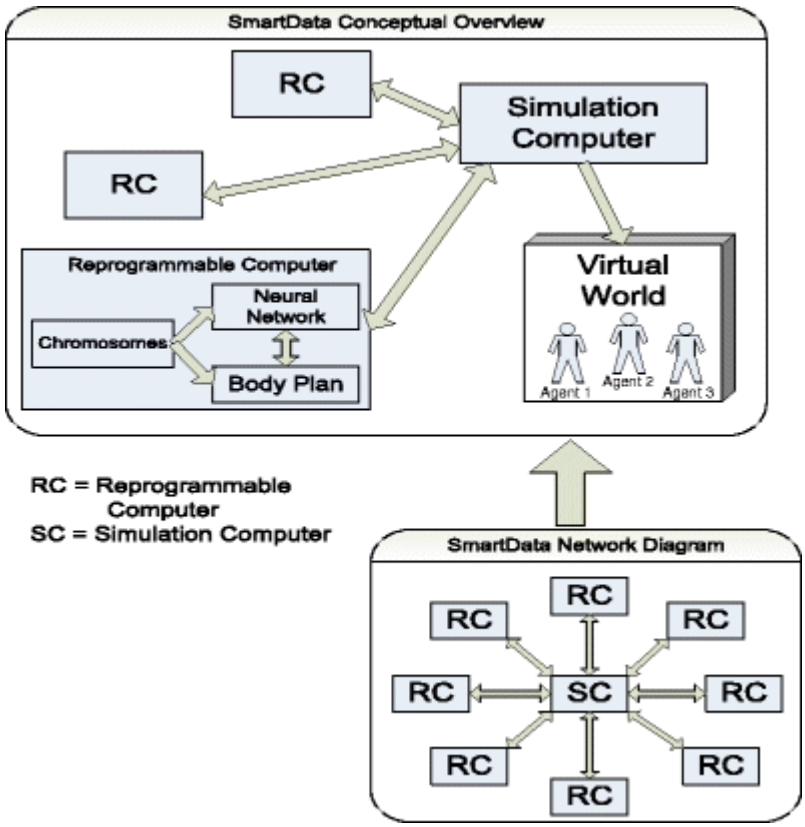
1. **Develop Algorithms for the Evolution of Cognitive Agents**
   This project involves the theoretical aspects of determining the appropriate evolutionary, learning and reasoning algorithms to enhance the likelihood of creating agents that are cognitive in nature and, most important for SmartData, also able to make decisions based on context. Evolution, learning, and reasoning are the three adaptive processes that operate at different timescales in promoting the survival of agents. Evolution, as the outermost loop, exploits long-term environmental dynamics that modify populations of agents. Learning, the middle loop, makes modifications to the neural network based on environmental dynamics within the “life of the agent.” Reasoning, the innermost loop, provides a mechanism for modifying behaviour consequent to future predictable variations in the environment[9]. They may be modeled as three selection procedures within an overarching evolutionary paradigm, each nested within another[10].
   It may be helpful to first provide some background on evolutionary algorithms in order to place this discussion into the proper context. Artificial evolution (like its natural counterpart), has an uncanny knack of producing powerful and efficient solutions to adaptive problems—solutions which the vast majority of human designers would not even contemplate. Evolutionary techniques are algorithms that search for solutions to a problem or task, within the space of all possible solutions. Theoretically, if the dimension of the space is large enough, a solution to any problem or task can be found. However, if the dimension is too large, the computation is intractable (even though evolutionary algorithms are amenable to parallel computing methods). Therefore, the art of using evolution is to limit the size of the solution space so that it is not too large, and yet may contain a satisfactory solution to the task. Nature gets around the dimensionality problem by using many parallel processors—that is, expanding the size of the population, where each individual may be viewed as a “processor.” But nature also has the luxury of time which allows it to explore many potential solutions and, bit by bit, select more adaptive ones.
   The methodology of evolutionary computation is to create a means of encoding “intelligent” control systems or agents as genotypes. Through this process, each agent is designed with chromosomes whose genes represent its “universe of behaviours” within its environment. The behaviour is produced via a neural controller acting on a “body.” A “body,” in our use of the term, is an interface to the world. It may range from a structure such as that possessed by humans in the real world, virtual agents or avatars in a virtual world or a simple digital code serving as the interface between two application programs. The agent’s genetic code specifies the range of behaviour that may be enacted by the controller, in concert with its body. The behaviour may be generated endogenously in the form of exploratory behavior or it may arise in response to an external stimulus.
   Starting with a population of agents with different genotypes (some may be random, while others will be designed) and an evaluation task which is built into the virtual world,[11] a selection cycle is implemented such that agents that are more successful in performing the task have a proportionately higher opportunity to contribute genetic material to subsequent generations, that is, to serve as future “parents.” Genetic operators analogous to recombination and mutation in natural reproduction are then applied to the parental genotypes to produce “children.” Each “individual” in the resulting new population is then evaluated to determine

how well it performed, and then the process is repeated, again and again. Over many successive generations, better performing agents are evolved.
Another way to think of the genotype or chromosome is as an input to a “cost function” which can be viewed as an evaluation task built into the world; the output of the cost function is the “fitness” (ability to survive) that a particular chromosome (specifying a particular agent) generates. The cost function can be a mathematical function, another algorithm such as a cryptographic one, an experiment such as a behavioural experiment, or a game. Cost functions may also be structured in hierarchical and nested fashion such that more complex behavior may be implemented. The benefit of genetic algorithms is that they are not necessarily captive to local maximums in the search space and can “hop out” via recombination and continue the search for a more global maximum. There are other techniques, such as those pioneered by the late Michael Conrad that we will explore called “extradimensional bypass” whereby increasing the dimensionality of the search space by one (i.e. adding an additional gene) can convert a local maximum into a saddle point thereby allowing continued “hill-climbing.”
As noted above, we are proposing a hybrid methodology, which we term “evolution by modifying design,” whereby populations of diverse simple agents are structured and programmed initially with different nested designs, in which the structural or algorithmic variables in these designs will be reflected in the agents’ chromosomes, and thereby modified with evolution. Some of these designs may be representative of “top-down” structures, as in classical AI, while others will parallel evolutionary algorithms representative of recent discoveries in Evolutionary Developmental Biology. These agents will be placed in a virtual evolutionary environment (a virtual Galapagos Island), with the objective that the “world” in conjunction with competition and cooperation from other agents, will select agents with chromosomal mutations that lead to higher levels of fitness—ones that are representative of SmartData. These techniques are standard practices in the field of evolutionary robotics and we plan to apply many of the same methods to evolution in a 3-dimentional virtual world.
Once the agents have evolved and learned a certain set of motor and social behaviours, the environment for SmartData will be modified to include:

1. Situations in which SmartData should allow its personal or proprietary information to be used for a primary purpose. These situations would be represented by other agents in the virtual world requesting the use of the information. The requesting agents could also be implemented in such a manner as to provide biometric samples (note that a biometric is usually defined as some measurable physical attribute that remains constant over time; we would select a similar attribute of an agent’s body).
2. Situations in which rogue agents attempt to access the information stored within SmartData but where SmartData should not allow access. These situations would include interception techniques such as spoofing, attempts to break the coding scheme, etc.
3. The situations described above would also be evolved as a population of rogue agents and placed in the environment of SmartData to form a competitive co-evolutionary environment. One of the benefits of evolving spoofing and code-breaking techniques is that methods that we humans have yet to think of could be anticipated and guarded against.

The initial design of agents that are placed in the virtual world for eventual evolution will include a priori rules to restrict use and place controls on the use of information, as set out by

the “surrogate” owner. These will function as constraints within which selected adaptive behaviours to a set of situations (stimuli built into the world) would be evaluated. This will allow the ingenuity of “genetic design” to find novel and ultimately more efficient solutions. This would also provide an additional security benefit since no one would know or be able to interpret how the rules and controls were implemented within the neural network. Thus, in effect, SmartData would become trusted software by virtue of its complexity. An interesting security method found in natural evolution, may be to incorporate “junk DNA” to make the process of breaking the code even more intractable. Clearly, there are many promising techniques for us to explore.



2. **Design of the Reprogrammable Computers (ReComs)**
The physical platforms or “body” for the agents will reside in the ReComs which will be located external to the virtual world and the Simulation Computer (“SimCom”). The ReComs will house the following components: the chromosomes consisting of the genes, the neural control system and the “body” plan of the agent. The genes will specify the nested structure of the neural control system, the rules for its operation and modification/learning capacities during its “life.” The genes will also specify the body plan, its degrees of freedom, actuator/movement capacities and initial connections between the neural control system and appendages/organs in the body plan. The body plan will be represented graphically in the virtual world as an agent/avatar by the SimCom, as discussed in Project 3 below. The genes will be mutated in accordance with the evolutionary algorithm. The mutated chromosomes will alter the internal structures of the agent housed within the ReComs. This will be carried out over a large number of cycles or generations. The specifications for SmartData will be embedded in the world as the fitness criteria which will ultimately select for survival; however, it will not be explicit as in standard genetic algorithms. It will actually be the “evolving” graphical avatar that will undergo

a process of selection by its environment, and should it survive, features of its specified neural control system and body plan, as represented by its chromosomes in the ReCom, will be retained, then recombined with those of another surviving agent, and passed on to its “offspring,” with a new cycle of mutations initiated.

One of the reasons we have chosen to have agents represented by distinct evolving hardware, external to the SimCom and its virtual world, is that, at the end of the day, when the evolutionary objective is reached and the agents possess the required “neural structure” for cognition, each agent will be represented by a string of binary digits which may then be moved around the Internet and downloaded into a similar ReCom—in effect, creating a clone of its previous self. This will also result in additional security benefits.

However, during the evolutionary stage of this project, each ReCom will have the capacity to house a number of different agents (‘n’ copies). While in the beginning, a ReCom will house only a single agent, should that agent survive, it will recombine its chromosomes with those of another surviving agent (initially from another ReCom but subsequently from within the same ReCom), and reproduce a select number of offspring. These offspring will reside within one ReCom. This procedure will allow populations of agents to grow without the need to introduce new hardware after each generation. Once the evolution is complete, however, the binary string representing the most successful individual agents may be copied/cloned into smaller versions of the ReCom hardware for SmartData applications.

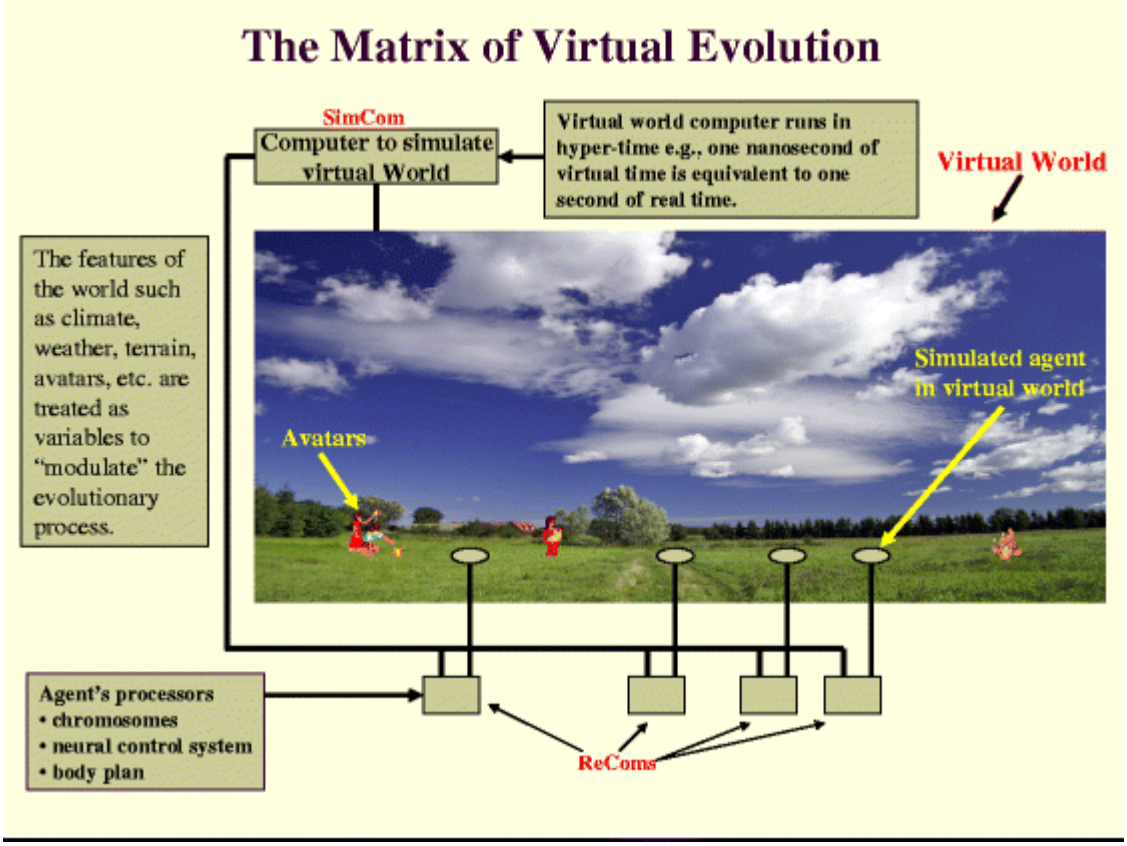
During the “life” of an agent, its neural control system will both receive and transmit signals to its virtual graphical avatar, based on stimulation from the virtual environment. We will also experiment with “signals” being received from its body plan within the ReCom to provide the status of its “body,” similar to internal messages received by one’s brain from one’s body. We consider this to be an essential dimension in the process of evolving agents with the capacity to take into account context. Although the body plan which is housed in the ReCom will be affected by changes to its associated virtual avatar, these changes will not be reflected in the genes themselves, as in Lamarckian evolution.

The development of reconfigurable hardware, ultimately suitable for this project, will require technology not yet available at a reasonable cost and will therefore be a major research direction. However, in the interim, we can use existing PC and acceleration hardware to simulate the operation of our ReComs. Our task in the ReCom project is to determine the best hardware and design specifications required to meet the challenge of malleable hardware that will be responsive to a changing virtual world.

3. **Design and Construction of the Virtual World**

This research program will be greatly facilitated by the use of an existing virtual world within which to evolve such agents. One example is the virtual world represented by Second Life. In such a world, agents would be able to interact with avatars which are “human-directed” and which provide a reasonably realistic social environment such that key properties that arise as a result of social interaction may be explored. However, an obstacle that will arise in such cases, is that in order to be “attention-grabbing” to the human-directed avatars already in the virtual world, the agents themselves would have to be embodied as “humanoid” or “mammalian,” possessing some type of agency, otherwise users in Second Life would not interact with them. The problem is that during the initial stages of any artificial evolutionary method, the agents will have random, “acausal agency” which will not be interesting to human-directed avatars. Furthermore, we do not believe that initially setting up these agents with a top-down algorithm or “programmed interesting behaviour” will lead to the cognitive or the “smart” property which is the objective of this research. In order to circumvent this problem, we therefore propose that initially, “dumb” non-humanoid agents be evolved within a separate “Galapagos Island” of

Second Life, until they develop a modicum of agency such that Second Lifers would be attracted to interact with them, which would in turn, stimulate further evolution towards SmartData.



Unlike most simulated evolutionary algorithms in which a fitness function incorporating behavioural objectives is stipulated, in the virtual world as in the real world, fitness cannot be measured directly but only assessed by observing the evolution of a population of agents. As designers, we must vary the parameters of the environments constituting the world and assess whether the resulting changes give rise to the desired behaviours. However, the world not only stimulates agents or organisms but selects them as well. It selects agents whose response patterns interpret stimuli in a beneficial manner. “Understanding what it is in the world that we respond to is at least as much a matter of understanding the ways in which the world selects cognitive mechanisms as it is of examining the behaviour of those mechanisms (Harms *2004*).” Selection by a dynamic world crafts the cognitive mechanism in which context serves as the differentiator to when a particular stimulus is viewed as beneficial. Accordingly, the environment in which agents are placed must be of ample complexity in order for evolution to select intelligent results. A simple world cannot select complex mechanisms. If we knew which features of the environment were responsible for selecting which properties of cognition, we could then construct an environment which only contained the features necessary to select the “cognition” required for the task. The problem, of course, is that we do not yet know which particular features, or combinations thereof, of the environment are necessary for the selection of specific properties in cognitive mechanisms. Therefore, it is imperative that we place the agents in a virtual world with attributes that parallel as closely as possible, the complexity of the world in which the agents will be required to operate. These attributes will vary based on the domain(s) in which the agents will operate.

The practices and procedures involved in safeguarding privacy and security on the Web are derivative from individuals’ concerns and solutions in the “real” world. These concerns and solutions are themselves derivative from the social and cultural environments in which we live. Therefore, if an agent in cyberspace is to function effectively it must first “understand” the specific social and cultural environment of humans in the area in which it will operate. An “electronic healthcare agent” does not need to understand the social and cultural environment of professional basketball; it must understand the environment within the domain of healthcare. Hence, the simulated virtual world must present the attributes of the relevant environment so that the proper cognitive characteristics will be selected for the job. This extends to the fact that SmartData agents must at some point in their evolutionary cycle inhabit a world with other agents in order to allow for inter-subjective cooperation and competition which, as has been demonstrated in our evolution, gives rise to particular social practices and cultures.

Although virtual worlds such as Second Life have approximated the “appearance” of the real world—improving continually as computer speed and graphical resolution improve, they have not yet incorporated a realistic physics engine. Although the initial experiments will be performed within the existing physics of a Second Life virtual world, we believe our research may eventually require a physics engine to simulate real-world phenomena such as gravity, laws of motion and thermodynamics (of which there already exists a number of open source and commercially available engines which may be incorporated into our virtual world). In addition, an “experience engine” will have to be built to first record the stimuli that the graphical avatars’ sensors experience in the virtual world and then transform those stimuli into appropriate signals which may be transmitted to the ReComs (for input to the neural network and body plan). The objective here is to program these factors as variables which may then be manipulated by researchers.

A factor which must be addressed in a later phase of our research is the speed of evolution. One of the many conditions that gave rise to cognition and its associated complexity in organisms was having the “right” amount of time available for evolution—approximately three billion years, from single-cell organisms to humans. The evolutionary paths that appear to have been successful were the result of low mutation rates over long periods of time[12]. In order to achieve interesting behaviours in our lifetime, we propose that we will have to “speed up evolution” significantly by simulating our virtual world in hyper-time. By that we mean, for example, representing one second of time in the “real” world as a fraction of a second in the virtual world. As an indication of the advantage of hyper-evolution, if in the future (when processor speeds are increased significantly from those achievable at the present time), we are able to represent one second of real time by one nanosecond of virtual time, this would allow for a million years of evolution to be carried out in 8 or 9 h in the virtual world. This will of course place great demands on not only the SimCom and ReComs but also the speed of our network. Initially though, we will begin with existing technology (which will limit the speed of evolution in the short term and the levels of complexity achievable).

4. **The Security of Personal Information Stored within an Agent**

This project will investigate the feasibility of structuring the security of data residing within SmartData into two lines of defense. As outlined previously, the data will first be divided into logical segments, with each segment encrypted using standard cryptographic techniques. Each segment will have a “non-personal” description of its data, as an attached meta-tag. The encrypted segments and associated plaintext meta-tag will then be embedded into the structure of a dynamical system as represented by the neural network within the ReComs. The objective of our project will be to demonstrate and verify that each agent will store this information in a differing fashion, as a result of possessing a different perspective from that of any other agent.

The structure of the neural network will change in a unique manner, as each agent evolves and mutates, based on its own experience of the virtual world. Accordingly, we believe there should be no single means of attack applicable to all agents. We propose that this property of possessing differing perspectives, in conjunction with a neural dynamical system, will provide greater security than traditional methods, and will seek to confirm that hypothesis.

5. **Networking Innovation Component**
The development network will be a star configuration in which a number of ReComs, each representing up to ‘n’ distinct agents, will be connected to the SimCom responsible for generating the virtual world. Each connection will need to emulate a “sensorimotor” cycle of about 100 ms for each agent housed within a ReCom: for example, once the sensors of a graphical avatar are “stimulated,” this information will be transmitted by the SimCom to ReCom’s neural network, and a motor response sent back via the SimCom to the graphical avatar for graphical action—the entire cycle occurring within about 100 ms.
Since there may potentially be a large number of “evolved” sensors within each agent and, as is the case of biological organisms, many sensors will be stimulated in parallel, there will be significant amounts of data that will need to be transmitted in a very short period of time between agents, embodied by the ReCom, and their graphically-represented avatar (within the virtual world simulated by the SimCom). Further, as the number of ReComs and agents increases, and both their world and agents become more complex, the amount of stimulus-response cycles, and in turn the bandwidths required, will increase dramatically. Thus, this program will require the development of both a high bandwidth network, presently available in the GENI project, and a sophisticated interface between ReComs and the SimCom.

References

Harms W. F. Information and meaning in evolutionary processes. Cambridge: Cambridge University Press; 2004.

Footnotes

1 See www.privacybydesign.ca

2 Newly developed methods of homomorphic encryption may alleviate some of these concerns.

3 http://cordis.europa.eu/fp7/dc/index.cfm?fuseaction=UserSite.FP7DetailsCallPage&call_id=219

4 http://www.opengroup.org/jericho/cloud_cube_model_v1.0.pdf

5 While for individual consumer records, the image format text may be manually “copyable” from a display or printer, we don’t envision this as a major problem due to the huge amounts of consumer data that would result; doing this en masse for an entire database would entail a very lengthy and cumbersome procedure in translating it back into digital format, for purposes of unauthorized use.

6 Although this may initially be time consuming for large databases of SmartData agents, once completed, the de-identified information could be used for a variety of purposes, without any further requests to the SmartData agents.

7 It could be no other way since in the virtual world in which they evolve and learn, no two agents can be at the exact same place, at the exact same time.

8 Using homomorphic or disk encryption methods may also enhance the security of SmartData.

9 In this report we use the terms “environment” and “world” interchangeably. The “world,” for our purposes, comprises two subsets: 1) the abiotic environment which is the climate, terrain, etc., which in our evolution changed relatively slowly, i.e., the changes were low in frequency; and 2) the biotic environment which comprises other interacting agents and species of agents, in other word, co-evolution, which has a higher frequency of associated change.

10 See Johnston, V. S., “Why we Feel” (1999) for a more thorough elaboration of the methodology of nested genetic algorithms.

11 By “built into the virtual world” we mean that the virtual world is structured such that in order to survive, specific behaviour(s) must evolve from a limited subset of behaviours. For example, humans placed in a frigid climate will have to evolve certain behaviours to survive. The subset of these behaviours may consist of discovering how to make fires to keep warm, hunt and use animal pelts as clothes to retain heat, build shelters, etc. The evaluation task is “built into the world” indirectly by setting the “thermostat” of the virtual climate to frigid temperatures. We do not specify a particular behaviour because that may limit the creativity and novelty of evolution to find solutions in the search space that we may not know even existed. Furthermore, nature is parsimonious in that it regularly selects behaviours and the associated genetic and neural structures that can serve as the basis for other behaviours built upon pre-existing ones.

12 In contrast, the chances of an organism at some stage in our ancestry withstanding a large number of simultaneous mutations (indicative of a high mutation rate) to its gene pool and still producing viable offspring, decreases as a function of the product of the probabilities of each mutation becoming beneficial. As a result, the time required to evolve interim viable organism under high mutation rates would be astronomically high (and we could still be at the single-cell stage). Therefore, the mutation rate has to be low enough to allow interim viability within reasonable times, which of course entails that there is enough time to serially compound these small viable mutations into complex structures and, in turn, behaviours.