

## **A Market for Secrets**

Eytan Adar and Bernardo A. Huberman  
HP Sand Hill Laboratory  
April 9, 2001

We propose an electronic market system for private data that guarantees levels of privacy, anonymity and control to individuals while maintaining the ability of other entities to mine their information and automatically pay individuals for their data. We also describe a novel procedure that allows data miners to anonymously contact the creators of the information in case their profiles are needed for future research.

Questions should be directed to [eytan@hpl.hp.com](mailto:eytan@hpl.hp.com) and [huberman@hpl.hp.com](mailto:huberman@hpl.hp.com)

# A Market for Secrets

## 1. Introduction

Privacy is an important and unresolved issue in the global society spawned by the Internet. The debate includes social, legal, and political issues that affect the ways through which individuals interact with companies, organizations and each other. The saliency of privacy on the Internet stems from the different and conflicting needs that users and providers have when it comes to the use of information. While individuals would like certain guarantees about how their data is obtained and used, few entities provide them. On the other hand, many institutions, both public and private, feel they have a right to access this data, whose analysis can also have positive effects for consumers, such as better products, cures for particular diseases, and even the creation of novel products that would satisfy consumer needs.

In spite of the stridency of the debate, it can be safely argued that most consumers would consider allowing access to some of their information if they were compensated in some form while their privacy would still be protected. An example of such a mechanism operating on a national scale is the current plan to establish a database containing the medical records of the entire population of Iceland, which has been made available to deCODE Genetics, a biotechnology company with the right to operate the database and sell access to third parties. Because of the unusual genetic homogeneity of the Icelandic population, such database could help identify the genes responsible for a number of diseases, which could in turn help design novel drugs for their treatment. Not only does the Icelandic population profit from such a deal on a financial basis, but also in the access to early diagnosis and treatment of their diseases.

In spite of the approval of this unusual deal by a majority of the Althingi, the Icelandic parliament, concerns are being voiced about the protection of the citizen's privacy. As has been pointed out[1-2], with a population of less than 300,000 people, a few pieces of data stripped of names, addresses and birth data can still reveal a person's identity, thus leading to fears that their privacy will be compromised. Nevertheless these concerns were not sufficient to prevent the deal with deCODE from proceeding, thus showing the value that Icelandic people put on access to their data.

It is with these considerations in mind that we propose an electronic market system for private data that allows consumers to receive automatic payments when others use portions of that data. Equally important, we introduce a novel mechanism, called *Information Crystals* that preserves anonymity and control on the part of the consumers while generating useful data profiles that companies and individuals can mine. This anonymity is accomplished through the digital analog of camouflage- the ability to blend into the surroundings- which we accomplish by allowing the information packets from individuals to hide amongst others with similar characteristics. Thus, while all the individuals contribute to a global picture that can be of use to individuals or firms, it becomes impossible to pinpoint the source of specific pieces of information.

*Information Crystals* is a distributed architectural framework that guarantees levels of privacy, anonymity, and control to individuals while maintaining the ability of other entities to mine their data and pay the users for their data. The information crystals are made up of atomic entities consisting of private records of data, which can range from genetic information and financial transactions to legal statements. Moreover, the size of these crystals guarantees that with high probability no single atom can be attributed to a single individual. There are a number of useful features that Information Crystals provide both to individual users and the data miners. Specifically:

- Users of Information Crystals can:
  - Create anonymous, private profiles that can be sold or given to data miners
  - Modify or delete their profiles
  - Have a guarantee that their identity is not compromised
  - Be compensated for the use of their information
- Data miners using Information Crystals:
  - Can hold user profiles locally
  - Obtain aggregate statistics for various user populations
  - While still being able to contact individuals for participation in further research or marketing offers

An important aspect to notice is that information crystals resolve the strong tension that has always existed between the privacy and anonymity of users and the demands of data miners and marketers. Generally, system designers must trade off some of the features described above. For example, how is it possible to be individually paid for the use of someone's data if that individual is anonymous? Likewise, how is it possible to guarantee users privacy if the data miners have a copy of the profiles?

While until recently these conflicting demands of users and data miners have been irreconcilable, recent developments facilitate the creation of an Information Crystal architecture. First, the cost of cryptography has dropped tremendously. New cryptographic accelerator boards from manufactures such as Broadcom, IBM, and nCipher [3] provide the necessary infrastructure to rapidly and easily perform cryptographic primitive operations. Second, the creation of encrypted mobile code [4] allows for the execution of code on remote, possibly malicious, hosts. And finally, recent developments in zero-knowledge protocols [5,6] allow for the matching of preferences between two entities without either knowing which specific items were matched. Thus, two users can conclude *how many* things they have in common, but not on *which ones* they match. Finally, AT&T's Crowds [7] architecture exemplifies a system by which individuals can hide their browsing patterns. Our system draws inspiration from the approach of collaboration among individual and applies it to pieces of mobile code to achieve global privacy goals.

In what follows we describe profiles as mobile code and some of the protocols and processes that make the Information Crystal system a powerful mechanism for private data mining. We also show how information atoms can be assembled into large crystals that can be queried at will. Finally we discuss mechanisms for implementing a market for secrets.

## 2. Profiles as Mobile Code

### 2.1 Information Atoms

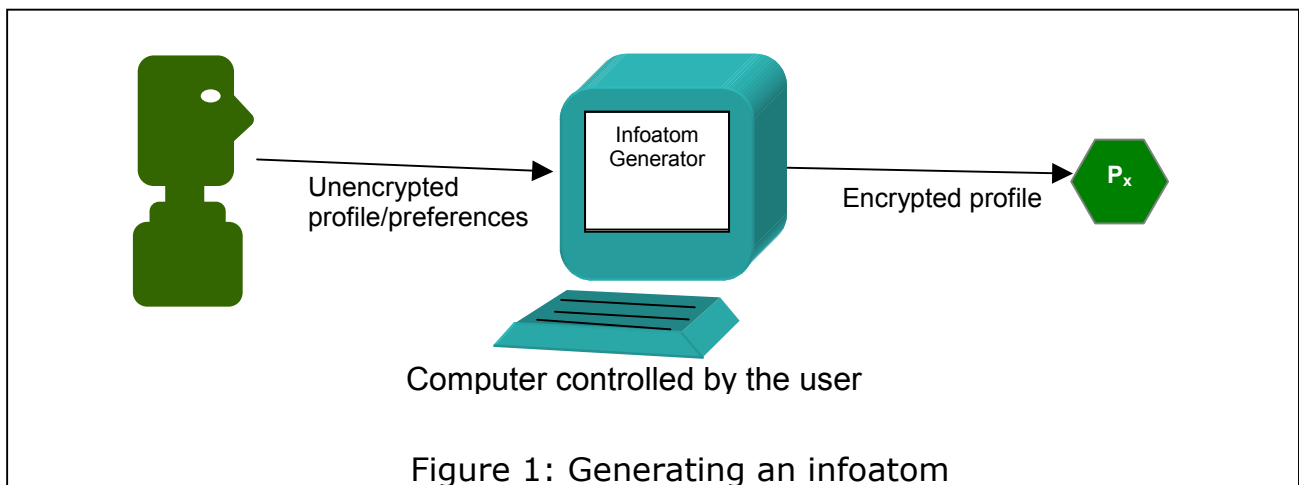
Information Crystals are composed of individual elements called Information Atoms (or infoatoms). Infoatoms are the encrypted, mobile, digital representations of an individual's profile and preferences. They also contain the keys that can be used to decrypt the data. Profiles can range from the address of the users and their preferences to their particular DNA sequence. Infoatoms are particularly powerful in that they contain a user's profile in encrypted fashion, as well as the computational logic necessary to interact with other infoatoms and the data miner while retaining the ability to function independently of the user that created them.

Figure 1 illustrates the creation of an information atom. Participants in this market, through the use of a program on their own computer, can encrypt their profiles and generate infoatoms. These infoatoms can then be distributed to data miners who might want to use the data for various purposes.

While infoatoms contain both the encrypted profile and the mechanisms for modifying, deleting, and revealing portions of this information, their most important feature is their ability to bind to other atoms so as to provide privacy and anonymity to the infoatom's creator.

### 2.2 From Information Atoms to Information Crystals

Given that infoatoms would like to hide among other atoms with similar properties it becomes vital to find those atoms. It is here that the zero-knowledge protocol for matching properties comes into play. Infoatoms use this protocol to determine overlaps between themselves and other infoatoms. The generation of information crystals through this bonding process is illustrated in Figure 2. Individual infoatoms (as shown in step 1) are activated by the data miner via some financial transaction. Once activated, an infoatom will randomly interact with other infoatoms and form connections (step 3). An infoatom will determine the number of other atoms with similar characteristics. If there are enough similar atoms the infoatom will become active (as seen in yellow). Once all infoatoms have bound to the crystal, the profile



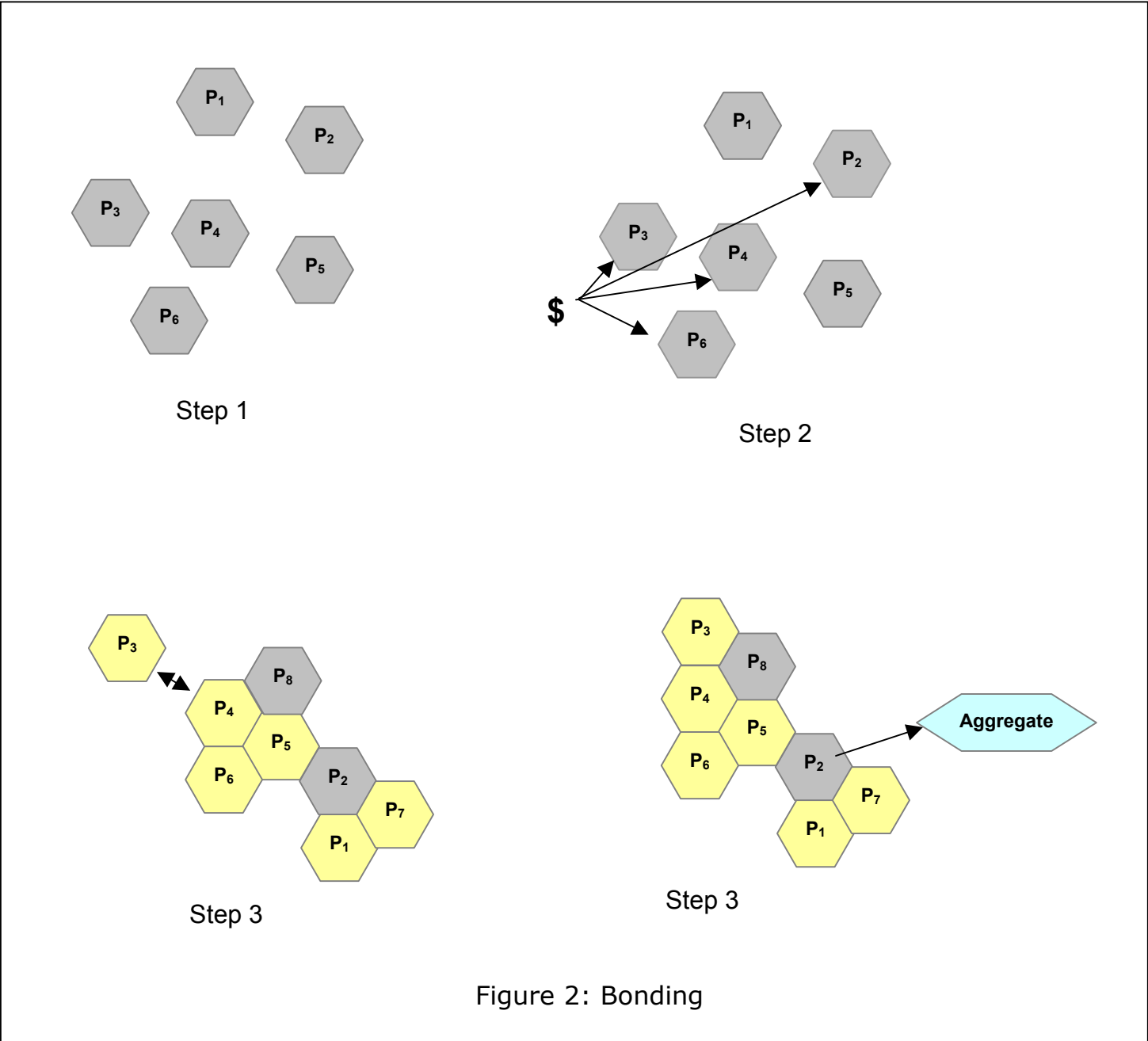


Figure 2: Bonding

data is emitted and collected by the crystal and the aggregate statistics are passed on to the data miner<sup>1</sup>.

**2.3 Querying the Information Crystal**

The protocol described above is a good way for obtaining total aggregate statistics for an infoatom population. However, this is of limited use to data miners who would like to have finer grained query control. For this purpose we introduce the notion of a *seed*. A seed is simply a special type infoatom that is generated by the data miner and which holds a query describing the miner’s desired output.

<sup>1</sup> The information crystal architecture provides the mechanism for anonymously generating this report, but the protocol specifics are beyond the scope of this paper.

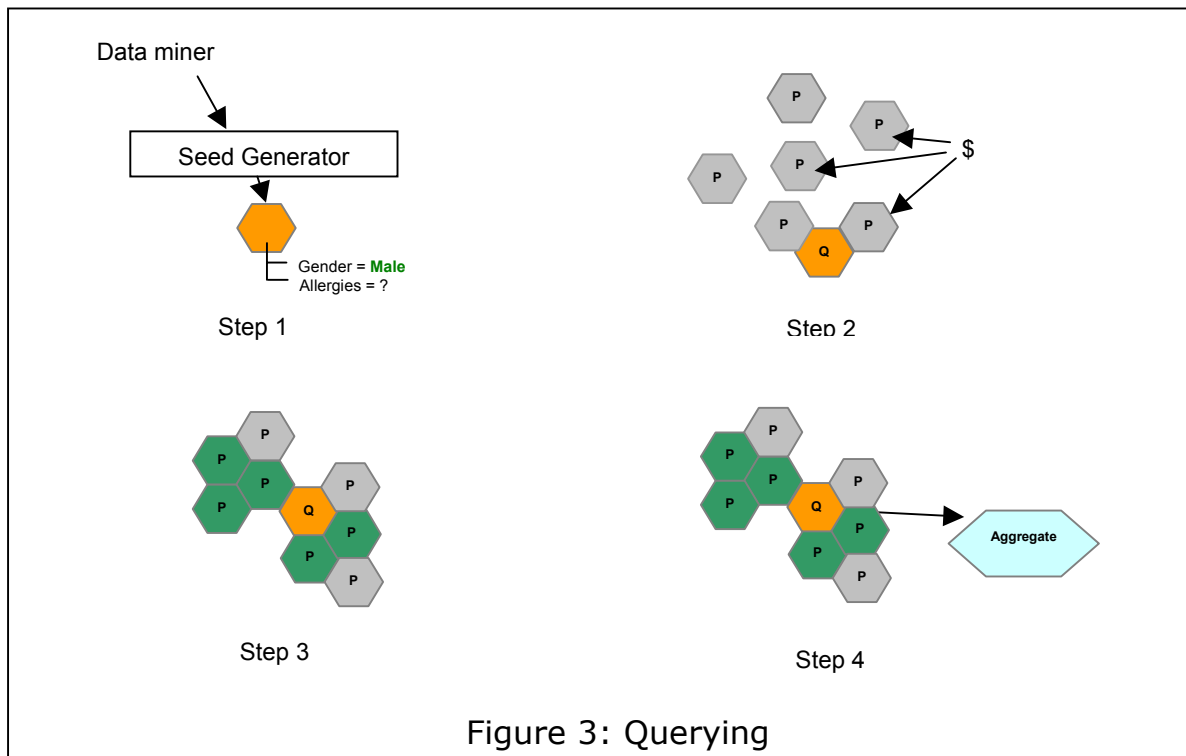
Figure 3 represents one such interaction between the data miner and an information crystal. In this scenario the data miner wishes to know particular allergies present in males with given characteristics and constructs a seed reflecting this query. The seed interacts with information atoms, which then cluster around the seed forming a crystal (step 2). Those atoms that match the query and that are surrounded by a sufficient number of similar infoatoms become activated in step 3 and eject an aggregate report in step 4.

## 2.4 Payments

One of the fundamental aspects of the Information Crystal architecture is the inclusion of an incentive mechanism through which users who have generated Infoatoms can be compensated for their use.

The architecture maintains this facility by incorporating into infoatoms the ability to accept micro-payments. There are a number of such mechanisms available that are capable of handling arbitrarily small amounts of money, while employing cryptography and Internet protocols to secure payments and automate risk management [8] As mentioned earlier, these micro-payments can activate the infoatoms and place them in a state receptive to bonding with other infoatoms so that they can release the aggregate information in response to a query.

An alternative approach to direct monetary compensation is the use of micro-stock payments. This would be suitable in situations where individuals who provide information to the data miner do not particularly understand or foresee the potential uses of the information they provide. If the company using the data is later to use the information in a product (say a pharmaceutical drug) both the individuals using the data and the company can benefit from the use of small stock grants. These



stock grants hold the promise of a valuation coupled to the success of the company (which partially depends on the successful use of the data).

One approach to making payments anonymous is to use an anonymous bulletin board protocols [9]. This allows the infoatom to publish the monetary token (the ability to collect on payments) encrypted with the users public key to a public bulletin board. The user collects payments by observing the contents of the bulletin board and decrypting those corresponding to his private key.

### **3. Security issues**

As with all systems that rely on notions of security and encryption, infomarkets as we described them can be vulnerable to some attacks. While not necessarily fatal, these attacks are important to describe, if only because countermeasures could perhaps be designed.

A trivial attack is one in which a data miner may infer infoatom properties by observing to which infocrystals that atom binds to. This can be countered by allowing infoatoms to randomly bind to non-matching infocrystals but to disclose "null" information in the aggregation step.

Another possibility would amount to a "disassembly" attack on the infocrystal itself. Since each infoatom contains the keys that can be used to decrypt its own data, someone could reverse engineer the infoatom to extract the keys and the encrypted profile, and then decrypt the profile data. This type of attack can be done either by a direct computational attack or what we term a "petri-dish" attack. A malicious individual could create many fake infoatoms with various preferences. By monitoring to which of these the real infoatom binds to and interacts with certain properties may be determined about the user. There are numerous strategies to counter such attacks. For example, one could break infoatoms apart into smaller structures where each maintains an unusable or uninteresting set of data, making the attacker unable to determine which infoatoms to apply the petri-dish attack to. Additionally, cracking each sub-atom can be made a costly process.

User privacy is not the only issue. It is possible that malicious users may create many infoatoms to either extract payment or corrupt aggregate statistics. This in some ways is the price of providing user anonymity. Such scenarios may be tempered by adding some expense to creating infoatoms or requiring third party authentication (such as a signature of the hospital that gave you your data). Additionally, if the incentives provided by the miner are catered to the properties of infoatom such that the rewards are unattractive to any use but those who actually have these properties (custom coupons, medications, etc.).

One final indirect issue is the inability of users to know what their data is used for. It may be possible, for example, to draw an inference between something seemingly unrelated such as "do you like the gym where you work?" (a question that seems harmless) to the probability of a heart attack. Additionally, individual users may not be identified as having a specific trait but the aggregate information may also result in discrimination. For example, a genetic data mining company could discover that members of a particular society (say Icelandic) are prone to acquire a particular disease. If a health insurance company were to have access to such data it could design a questionnaire for prospective customers so as to elicit nationality even by

indirect means and rejects or increase cost to those of Icelandic origin. Because it is difficult to predict such uses, it may be necessary to address these issues through legal, educational, and economic institutions in addition to technological means.

#### **4. Conclusion**

In this paper we proposed an electronic market system for private data that allows consumers to receive automatic payments when others use portions of that data. Equally important, we introduced a novel mechanism, called *Information Crystals* that preserves anonymity and control on the part of the consumers while generating useful data profiles that companies and individuals can mine. This offers a resolution to the strong tension that has always existed between the privacy and anonymity of users and the demands of data miners and marketers.

Several authors have expressed a need for such a mechanism in various fields, including marketing and health and genetic databases but as yet no one has proposed a solution that requires no trusted third parties while preserving individual privacy. We believe that the coupling of a market for secrets with the information crystal architecture we described goes a long way towards providing such a solution.

#### **5. Acknowledgements**

We thank Matt Franklin, Rajan Lukose, Patrick Scaglia, and Jaap Suermondt for useful comments and discussions. Part of this work was initiated while the authors were at Xerox PARC.

#### **6. References**

- [1] Martin Enserink, "Start-Up Claims Piece of Iceland's Gene Pie," *Science* 2000 February 11; 287: 951
- [2] R. Chadwick and Kare Berg , "Solidarity and Equity: new ethical frameworks for genetic databases," *Nature Genetics* Vol. 2, 318 (2001)
- [3] Lori MacVittie, "Cryptographic Accelerators Provide Quick Encryption," *Network Computing*, April 19, 1999.
- [4] Christian Cachin, Jan Camenisch, Joe Kilian, and Joy Müller. One-round secure computation and secure autonomous mobile agents. In Ugo Montanari, José P. Rolim, and Emo Welzl, editors, *Proc. 27th International Colloquium on Automata, Languages and Programming (ICALP)*, volume 1853 of *Lecture Notes in Computer Science*, pages 512-523
- [5] Bernardo A. Huberman, Matthew Franklin and Tad Hogg, "Enhancing Privacy and Trust in Electronic Communities", *Proceedings of the ACM Conference on Electronic Commerce*, 78-80 (1999).
- [6] M. Naor and B. Pinkas, "Oblivious polynomial evaluation", available at <http://www.wisdom.weizmann.ac.il/~naor/onpub.html>.

[7] "Crowds: Anonymity Loves Company," AT&T Research Laboratory, various papers at: <http://www.research.att.com/projects/crowds/>

[8] For a micropayments overview, see <http://www.w3.org/ECommerce/Micropayments/>

[9] See for example, F. Stajano and R. Anderson, "The Cocaine Auction Protocol: On the Power of Anonymous Broadcast" , 3<sup>rd</sup> International Workshop on Information Hiding, Lecture Notes in Computer Science, Springer, (2000).