Response to *Request for Input (RFI) — National Big Data R&D Initiative*

Contact Information:

Dr. W. Knox Carey, Vice President, Technology Initiatives, Genecloud/Intertrust Technologies Corporation, 920 Stewart Drive, Sunnyvale, CA 94085. Telephone: (408) 616-1666. Email: knox@intertrust.com.

Our Role in the Big Data Ecosystem

The Genecloud project is an initiative of Intertrust Technologies Corporation, a private company in California focused on technologies for preserving individual privacy in the analysis of big data. The company is engaged in a number of technical initiatives in various application areas, including behavioral targeting for advertising, privacy for automotive data, home energy monitoring data, and others — areas in which analytics from multiple parties are essential for providing services and improving quality of life, but also areas in which compromise of personal data can cause great harm.

Healthcare information is arguably among the most sensitive personal information in this category. Our Genecloud project is focused on finding a pragmatic balance between access to healthcare data for analysis and patient privacy. Research in this project builds on work in trusted distributed computing, cryptography, policy management, information flow control, bioinformatics, and many other disciplines.

Comments on *Visions and Priority Actions*

We believe that the vision needs to be expanded to encompass the rights of individuals. Every datum in a Big Dataset was collected from a person whose rights must be respected — the right to remain anonymous, the right to negotiate the terms under which private data are used, the right to withdraw information from a data set. For reasons of expediency these issues are often ignored or traded away, but technical solutions exist that allow the parties that rely on big data collected from individuals to solicit and enforce their consent for the intended uses.

These techniques are needed particularly when multiple agencies are involved in data analysis. While individuals may be willing to consent to certain uses of their private information by a first agency, this trust does not necessarily extend transitively to additional agencies, no matter how trusted by the first.

In addition, we believe that the second vision point (i.e. *to understand trustworthiness of data and resulting knowledge*) is essential. Analyses — and especially analyses that are made by a series of collaborating entities — should be accompanied by a *digital chain of handling* that allows all relying parties to understand precisely how, when, and by whom results were derived. Ideally, this manifest would also include code, initialization parameters, and other information that would aid in scientific reproducibility and *post facto* validation of results. It is a primary goal of our research to develop such tools.

In terms of research, education, and infrastructure priorities, we would note that very often issues of privacy and data security are treated as secondary priorities, to be addressed once an initial research goal is achieved. However, because security is a cross-cutting systemic concern, retrofitting data protection, trust management, and policy to an existing system seldom succeeds. The Internet itself is a case in point: the retroactive application of security into systems such as DNS, SMTP, and HTTP have been difficult and error-prone. To design more resilient systems, designs must incorporate specific security and privacy goals from the start, and this begins with education and infrastructure investments.

The federal government should lead by example on these issues. Private companies whose businesses depend on data collected from individuals have often neglected privacy — and in some cases acted explicitly against it — in their quest for larger data sets. In its inter-agency programs and in joint governmental/commercial projects, the federal government should insist that security and privacy controls be in place as a condition for funding or engaging in any collaboration. In addition, the federal government should set research and funding priorities to further these goals.

Privacy is a Fundamental Component of a Strategic Plan on Big Data

Big Data technologies will fundamentally change society, ushering in innovations that will improve our lives in ways that we can scarcely imagine. In the rush to adopt these technologies, however, we must be cognizant of their effects on individual privacy.

In healthcare, in particular, we must not be asked to choose between two equally fundamental rights: the right to lifesaving information and the right to personal privacy. The technologies needed to preserve and manage privacy do exist, and need not interfere with the analysis and sharing of big data. Indeed, we would argue that the presence of privacy guarantees will actually *increase* data sharing by mitigating many of the most serious risks.

We respectfully request that the National Big Data R&D Initiative incorporate research, educational programs, research goals, guidelines, and explicit requirements around individual privacy and policy control into its strategic plan. Our organization is committed to these goals and stands ready to assist in any capacity.