

Statement of Latanya Sweeney, PhD

Visiting Faculty, Harvard University and MIT;

Distinguished Career Professor of Computer Science, Technology and Policy  
and Director of the Data Privacy Lab, Carnegie Mellon University;

GAO Appointee to the Privacy and Security seat of the Federal HIT Policy Committee

before the 21st Century Healthcare Caucus Roundtable

“Designing a Trustworthy Nationwide Health Information Network (NHIN) Promises Americans Privacy and Utility, Rather than Falsely Choosing Between Privacy or Utility”

April 22, 2010

Representative Murphy, Representative Kennedy and other respected members of Congress, thank you for the opportunity to testify today.

Given HIPAA [1], meaningful uses [2], and the current NHIN (Nationwide Health Information Network) [3],

I will show that HIPAA alone is not sufficient to protect patients from harm. New kinds of harms can result from the design of the NHIN. Rather than new policy however, we propose a risk analysis of NHIN design patterns to identify sweet spots for best achieving privacy and utility through technology and policy integration. We hope you will encourage ONC (Office of the National Coordinator) to support our performing this risk analysis.

Since the passage of HIPAA, there has been an explosion in the collection and sharing of patient information. While HIPAA explicitly identifies covered entities that handle patient information, there is no identification of the vast number of business associates who receive patient information from covered entities, or of the business associates of those business associates, and so on, as secondary sharing is unbounded. Data sharing through business associate arrangements is widespread yet hidden from patients, making harms difficult to trace.<sup>1</sup>

As data sharing increases, the number of autonomous copies of patient information increases, making controls difficult to enforce. Figure 1(a) appeared in a National Research Council report in 1997 [4]. It describes data sharing of patient information before the promulgation of the HIPAA Privacy Rule. Figure 1(b) depicts the greater number of copies today [5], which is before widespread EMR adoption and the NHIN.

Here is another example of the increase in data sharing. In 2002, two companies licensed my risk assessment technology in order to provide HIPAA certifications for de-identified data. A typical request in 2002 contained about 15 fields of information from single encounters of a few thousand patients in a specific geographical area, with the request made by a HIPAA covered entity. Today, the typical request has about 300 fields of linked longitudinal encounters of millions of Americans from around the country, with the request made by an organization downstream of a HIPAA covered entity.

---

<sup>1</sup> We have an integrated technology and policy approach to provide more transparency in secondary data sharing under HIPAA, but discussion of such lies outside this specific testimony. We do welcome a future opportunity to discuss this problems and the space of possible remedies.

Having meaningful users of EMRs (electronic medical record systems) as encouraged by ARRA [6] will further increase data sharing, but the most dramatic increase will be a consequence of benefits made possible by nationally sharing patient information over the NHIN.

The NHIN promises many benefits. Perhaps the most visible benefit to American patients is the ability to have allergies and medications provided at the time and place of service. If while visiting DC, I became unconscious and presented in a DC emergency room, the NHIN would allow attending physicians to reference my medications and allergies from Pennsylvania.

I have described massive data sharing on one hand and improved patient care on the other. These underscore the tension between privacy and utility. Reaction to this tension often harbors a false belief that one must be traded against the other –i.e., in order for society to reap the benefits of the NHIN, Americans must give up privacy. Unfortunately, the current approach to NHIN design is worse, making it unlikely Americans will have either privacy or utility.

Figure 2(a) depicts the traditional false belief of trading privacy for utility. It also shows our 9-year experience in the Data Privacy Lab of finding sweet spots of solutions that give privacy and utility. The key to our success has been technology design.

The current approach to NHIN design can be characterized as “let a 1000 flowers bloom.” Or, as someone close to the operation said, “let 1000 weeds fester.” The lack of architectural direction allows simultaneous efforts to proceed in different, even opposing directions, exposing patient information to various risks and limiting benefits. States and regional organizations are making independent isolated decisions. Various competing industry efforts are underway. And, national efforts recognized by ONC are inconsistent and problematical.

For example, ONC’s website describes the NHIN Limited Production Exchange as today’s NHIN [7]. This effort is driven by federal entities (e.g., the Social Security Administration, the Veteran’s Administration, the Department of Defense, and the CDC) with participation from private entities (e.g., Kaiser Permanente).

Technical operations are: (1) *patient lookup*; (2) *document query*; (3) *document retrieval*; (4) *audit log query*; (5) *authorized case follow-up*; and, (6) *event messaging* [8]. These functions allow participants to locate patient information, identify available documents, retrieve patient documents, etc. The last operation, *event messaging*, allows criteria to be stored so that when a patient appears matching the criteria, the patient’s information automatically forwards. Overall, this is not a bad set of operations, but there are many serious problems in its design. Here are a few.

Given explicit identifiers of a patient (e.g., name and SSN), *patient lookup* returns institution-specific identifiers for that patient [9]. Using a network of patient registries, where each register contains each patient’s name, telephone, gender, date of birth, Social Security number, and optionally, deceased information, marital status, religious affiliation, race, ethnicity and address, along with the patient’s unique identifier at each organization. There are numerous concerns with this approach. One concern is that insiders can do malicious surveys to locate patients and patient records. For example, a domestic violence stalker can use the system to locate victims.

In one version [10], *event messaging* allowed 3<sup>rd</sup> party notification of patient information outside the direct care of the patient and without the patient’s knowledge. This function may help public health agencies receive information on reportable conditions in their jurisdictions, but designed in this

manner, it enables unsupervised surveillance. For example, an insider could receive notifications of all abortions performed at other organizations.

To participate, institutions maintain portals accessible to patient information; and if used widely, these portals would be in environments (e.g. provider groups) having no professional staff to follow the latest computer security practices<sup>2</sup>.

(There are many other issues with this design too.)

ONC's website also describes NHIN Direct [11] as a parallel initiative underway [3]. The idea came from comments made by representatives from Microsoft and Cerner [12]. In current practice, two providers fax patient information as needed. So, the idea is to replace the fax with email that has secure channels to combat eavesdropping. There are numerous concerns with this design also. A glaring problem is its limitation. We cannot perform all meaningful uses with this system, so we will need an additional system, which begs the question: why build this system at all? For example, this design cannot reasonably retrieve allergies and medications for an unconscious patient presenting at an out-of-state emergency room (arguably a stage 1 meaningful use).

Figure 2(b) summarizes concerns about these two designs. The NHIN Limited Production Exchange has serious privacy issues but more utility than NHIN Direct. On the other hand, NHIN Direct has fewer privacy issues, but insufficient utility. When combined, we realize the least of each design, providing an NHIN with limited utility and privacy concerns.

Moving beyond these specific designs, we examine a large array of trust issues any NHIN design must address to be widely accepted. Here are some examples.

De-duplication and identity (see Figure 3). A useful goal of an NHIN is to get a distinct global count of patients matching a given criteria. For example, public health may want to know the number of Americans diagnosed with H1N1. The common approach is for providers to share patient identifiers (e.g., patient SSNs) rather than providing counts of cases at their facilities so the recipient can perform de-duplication (count each patient once even if the patient visits multiple reporting providers). The widespread sharing of identifiable patient information generates a privacy concern.

Testing and liability (see Figure 4). An expected source of cost savings attributed to health data sharing is an anticipated reduction in duplicated tests and procedures. The goal is to reuse test results. So, we think in terms of making sure the source is reputable, the image is not modified, and the information is accurate and belongs to the correct patient. (Technically, we may think of digital signatures.) But providers are concerned about liability. Credentials are not the same as having a trustworthy relationship with a lab or removing malpractice risk.

Corrections (see Figure 5). In the data sharing environments described so far, there is no mechanism for propagating corrections or updating patient information.

Data segmentation (see Figure 6). The idea of data segmentation is to provide patients a form of privacy by sequestering some data. Attempting to do so usually generates an inference problem, where

---

<sup>2</sup> If these systems must be certified by FDA, the machines may not be eligible for timely security patches, as the patches themselves may have to go through a certification process before installing. In the interim, machines would be vulnerable to known attacks.

the missing information can be learned from what remains. In health data, the inference problem is compounded because information is replicated in autonomous copies of information held in different organizations.

Care and Safety (see Figure 7). An overall goal of the NHIN is for providers to use consolidated information over time, compiled from different providers to improve decision-making for the patient. Lack of trust by patients or providers will result in missing or incorrect data, and will thereby undermine the entire effort.

This is small sample of the trust issues that any NHIN design must address to ensure stakeholder adoption. So, our approach is not to construct a single design, but to provide knowledge about the kinds of issues and possible solutions in design patterns that may appear in the “weeds” under development. We envision producing a spreadsheet where each row is a design pattern and each column is a property or issue important to NHIN operation. We expect each design pattern to have problems, but some designs will have more and others less, and the nature and space of possible remedies will be different. Having this information makes it easier to compare designs and to understand what kinds of accompanying policies are necessary.

Figures 8 through 13 illustrate some of the design patterns we would like to analyze. Global query (Figure 8) answers queries across a collection of repositories in real-time like web searching. Certified Delivery (Figure 9) provides point-to-point delivery of information with verifiable and accountable endorsements at the sender and receiver ends. Patient Central (Figure 10) uses globally available storage of patient information arranged by patient, analogous to a virtual file cabinet. Medical Billing Backbone (Figure 11) leverages the existing connectivity of providers to the billing system. Heavy Data (Figure 12) tethers meta information to data values, maintaining data provenance. Pointer Addressing (Figure 13) shares storage addresses rather than the values themselves, making corrections easier. While there are other design patterns we consider, I highlighted this group because each of these seem well suited to address at least one of the trust issues described. No one design patterns appears best for all issues.

The harshness of my comments is not aimed at ONC or its people. I have voiced concerns over the NHIN design since last summer and ONC has been responsive, but those efforts (new hires and working groups) have not yielded the best results, because success is best realized by tight integration of technology and policy and the technology is immature. As Aneesh Chopra said, “We are not building on a firm legacy of success, we are looking for a pathway to success.” [13] Performing risk analyses on design patterns provides a clear, informative path.

Thank you.

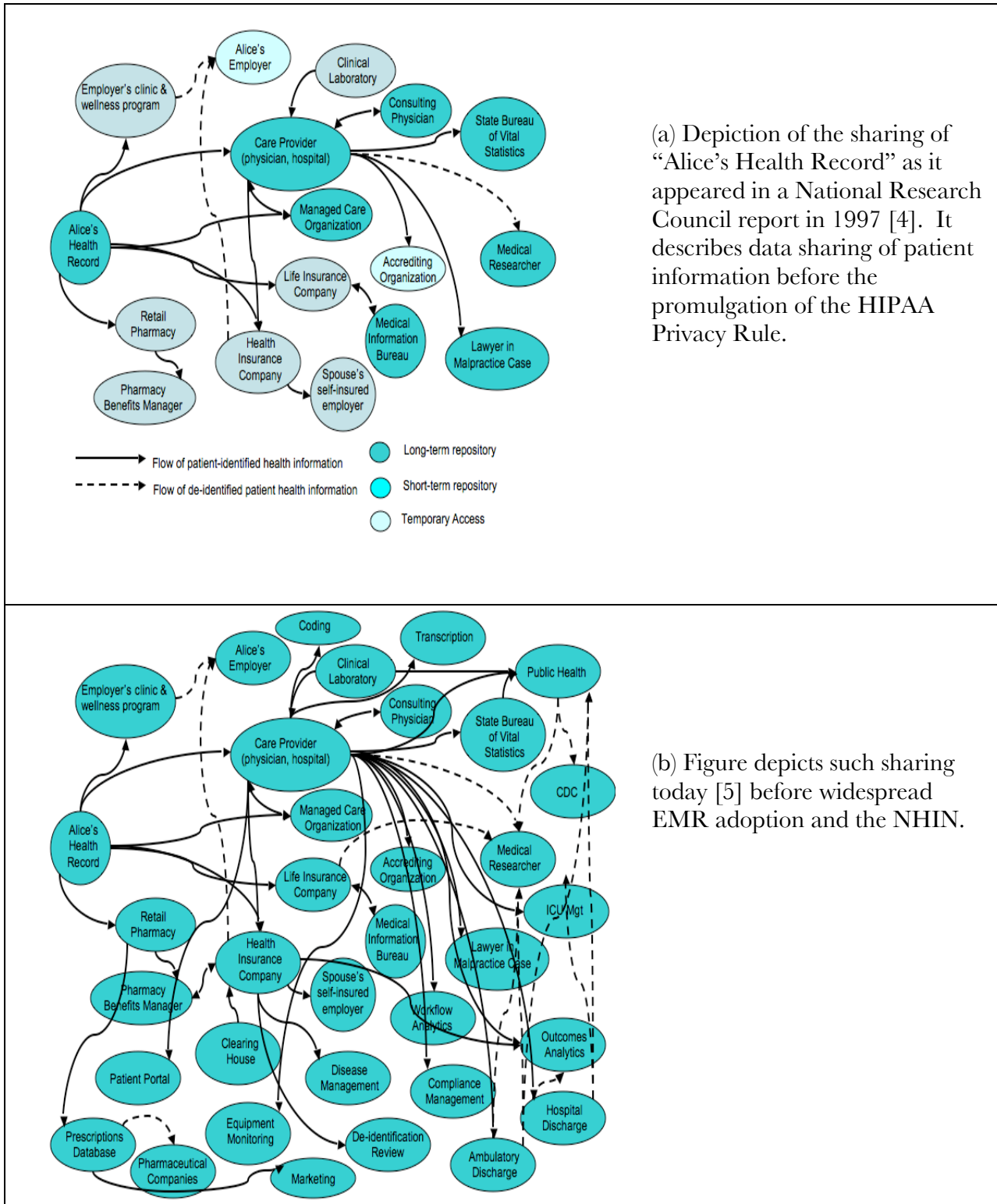
Latanya Sweeney, PhD

latanya@seas.harvard.edu  
dataprivacylab.org/people/sweeney

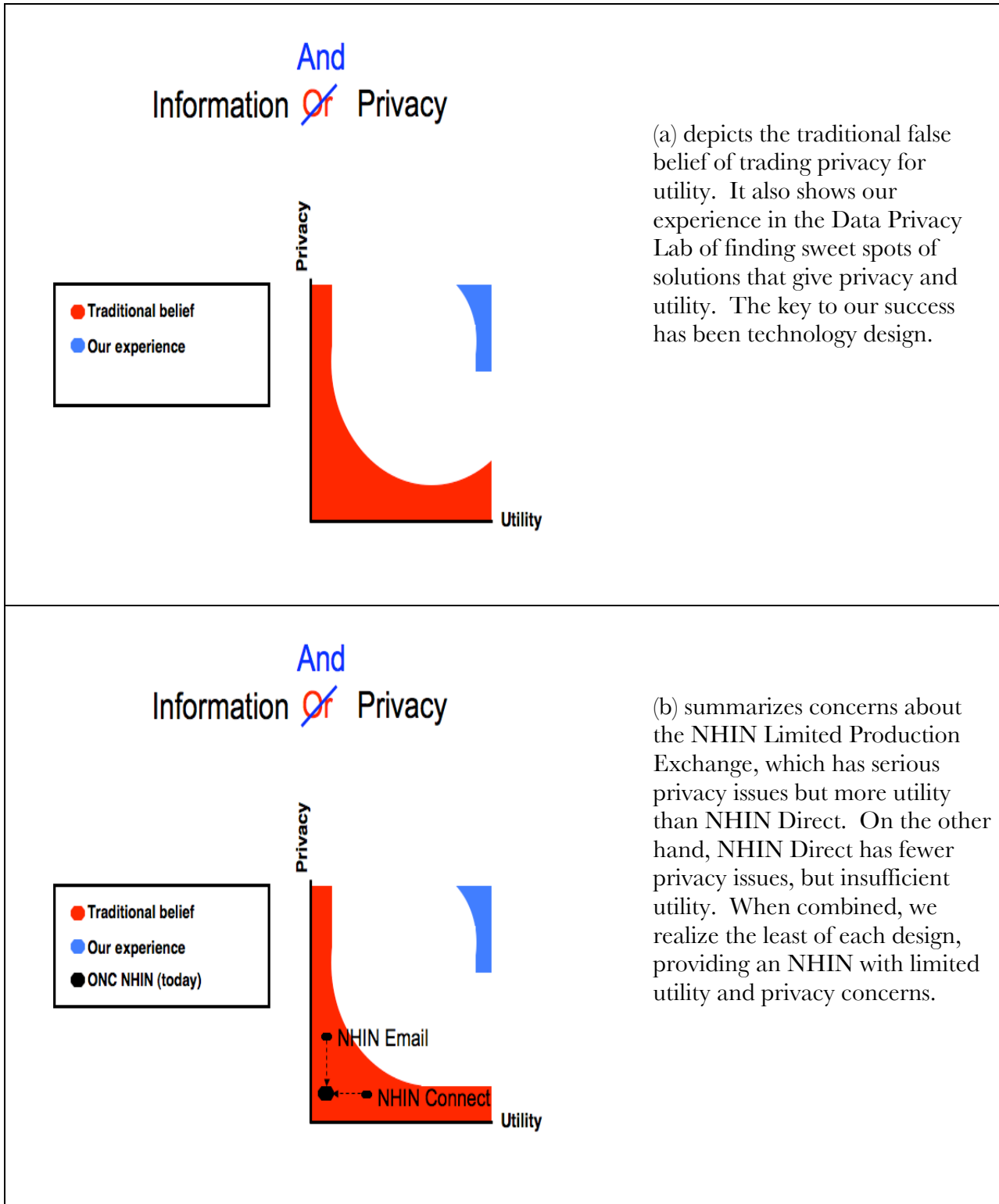
**References**

- 1 Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy Rule. 45 CFR Parts 160 and 164. As of 4/21/2010 <http://www.dhhs.gov/ocr/privacy/hipaa/administrative/privacyrule/index.html>
- 2 42 CFR Parts 412, et al. Medicare and Medicaid Programs; Electronic Health Record Incentive Program; Proposed Rule. As of 4/21/2010 <http://edocket.access.gpo.gov/2010/E9-31217.htm>
- 3 Nationwide Health Information Network (NHIN). Office of the National Coordinator. U.S. Department of Health and Human Services. As of 4/21/2010 [http://healthit.hhs.gov/portal/server.pt?open=512&objID=1142&parentname=CommunityPage&parentid=25&mode=2&in\\_hi\\_userid=11113&cached=true](http://healthit.hhs.gov/portal/server.pt?open=512&objID=1142&parentname=CommunityPage&parentid=25&mode=2&in_hi_userid=11113&cached=true) (Archived at [advancehit.org/NHINarchive/1/](http://advancehit.org/NHINarchive/1/))
- 4 Clayton, P. et al. For the Record: Protecting Electronic Health Information. National Research Council. National Academy Press. 2007. As of 4/21/2010 [http://www.nap.edu/openbook.php?record\\_id=5595](http://www.nap.edu/openbook.php?record_id=5595)
- 5 Sweeney, L. Alice's Electronic Health Information Revisited. Carnegie Mellon University. AdvanceHIT Project. Working Paper 1004. December 2009. <http://advancehit.org/publications/p1004/index.html>
- 6 American Recovery and Reinvestment Act of 2009 (ARRA). Public Law 111 – 5. <http://www.gpo.gov/fdsys/pkg/PLAW-111publ5/content-detail.html>
- 7 NHIN Limited Production Exchange. Office of the National Coordinator. U.S. Department of Health and Human Services. As of 4/21/2010 [http://healthit.hhs.gov/portal/server.pt?open=512&objID=1407&parentname=CommunityPage&parentid=8&mode=2&in\\_hi\\_userid=11113&cached=true](http://healthit.hhs.gov/portal/server.pt?open=512&objID=1407&parentname=CommunityPage&parentid=8&mode=2&in_hi_userid=11113&cached=true) (Archived at [advancehit.org/NHINarchive/2/](http://advancehit.org/NHINarchive/2/))
- 8 2010 NHIN Final Production Specifications. Office of the National Coordinator. U.S. Department of Health and Human Services. As of 4/21/2010 [http://healthit.hhs.gov/portal/server.pt?open=512&objID=1407&parentname=CommunityPage&parentid=8&mode=2&in\\_hi\\_userid=11113&cached=true](http://healthit.hhs.gov/portal/server.pt?open=512&objID=1407&parentname=CommunityPage&parentid=8&mode=2&in_hi_userid=11113&cached=true) (Archived at [advancehit.org/NHINarchive/3/](http://advancehit.org/NHINarchive/3/))
- 9 NHIN Patient Discovery Web Service Interface Specification v1.0. Office of the National Coordinator. U.S. Department of Health and Human Services. As of 4/21/2010 [http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS\\_0\\_11673\\_910524\\_0\\_0\\_18/NHIN\\_PatientDiscoveryProductionSpecification\\_v1.0.pdf](http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_11673_910524_0_0_18/NHIN_PatientDiscoveryProductionSpecification_v1.0.pdf) (Archived at [advancehit.org/NHINarchive/3/](http://advancehit.org/NHINarchive/3/))
- 10 NHIN Connect. As of January 25, 2010, at <http://www.connectopensource.org/display/Gateway/CONNECT+Community+Portal> (Archived at [advancehit.org/NHINarchive/4/](http://advancehit.org/NHINarchive/4/))
- 11 NHIN Direct. As of 4/21/2010 <http://nhindirect.org/> (Archived at [advancehit.org/NHINarchive/5/](http://advancehit.org/NHINarchive/5/))
- 12 Nolan, Sean. Prepared Remarks on Authentication. Testimony before the HIT Policy Committee NHIN Workgroup. January 7, 2010.
- 13 Chopra, A. Briefing from the HIT Standards Committee, Implementation Workgroup. Testimony before the HIT Policy Committee 12/15/2009. [http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS\\_0\\_11673\\_909821\\_0\\_0\\_18/ChopraImplementationWG121509.ppt](http://healthit.hhs.gov/portal/server.pt/gateway/PTARGS_0_11673_909821_0_0_18/ChopraImplementationWG121509.ppt)

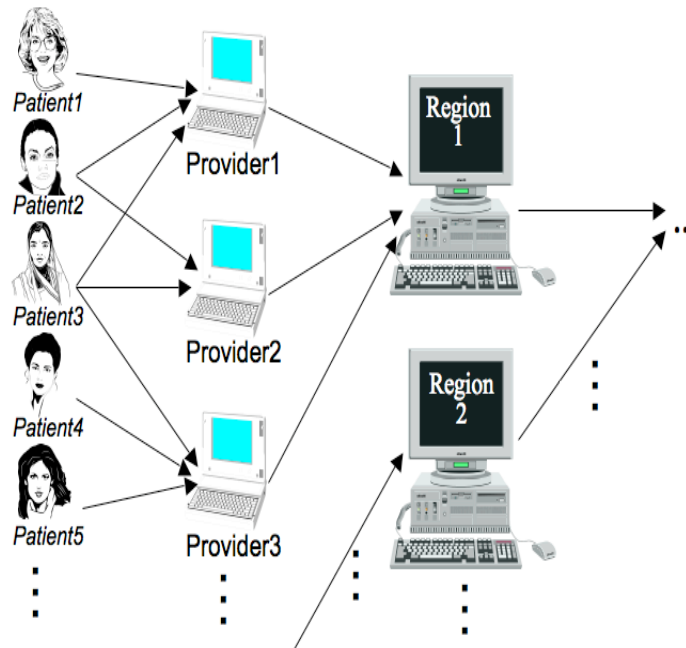
**Figure 1. Depiction of the sharing of “Alice’s Health Record”**



**Figure 2. Information and/or Privacy Space of Solutions.**



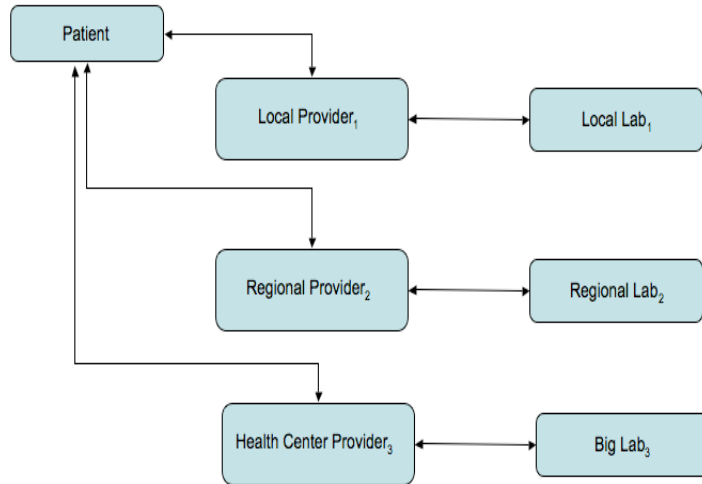
**Figure 3. De-duplication and Identity (Example of an NHIN Trust Issue)**



A useful goal of an NHIN is to get a distinct global count of patients matching a given criteria. For example, public health may want to know the number of Americans diagnosed with H1N1. The common approach is for providers to share patient identifiers (e.g., patient SSNs) rather than providing counts of cases at their facilities so the recipient can perform de-duplication (count each patient once even if the patient visits multiple reporting providers). The widespread sharing of identifiable patient information generates a privacy concern.

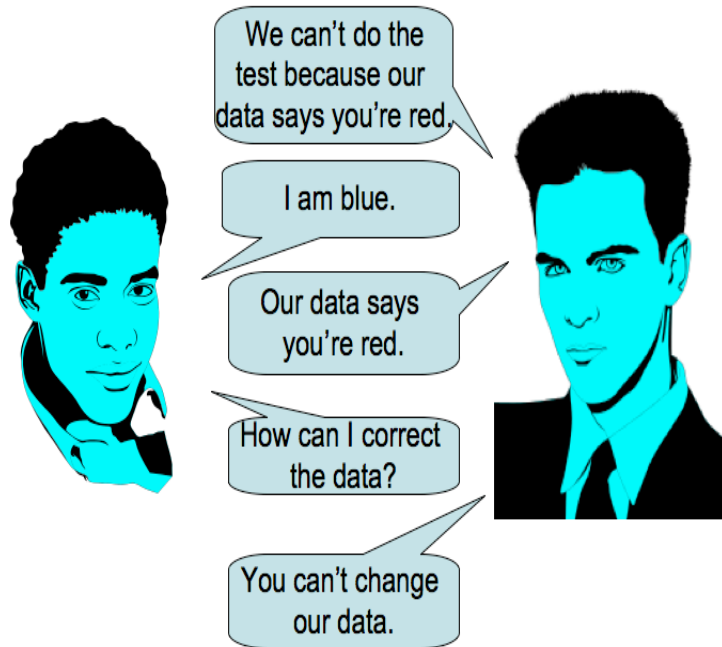


**Figure 4. Liability and Testing (Example of an NHIN Trust Issue)**



Providers trust the labs with which they work regularly. An expected source of cost savings attributed to health data sharing is an anticipated reduction in duplicated tests and procedures. The goal is to reuse test results. So, we think in terms of making sure the source is reputable, the image is not modified, and the information is accurate and belongs to the correct patient. (Technically, we may think of digital signatures.) But providers are concerned about liability. Credentials are not the same as having a trustworthy relationship with a lab or removing malpractice risk.

**Figure 5. Corrections (Example of an NHIN Trust Issue)**



In the data sharing environments described so far, there is no mechanism for propagating corrections or updating patient information.

**Figure 6. Data Segmentation (Example of an NHIN Trust Issue)**

Lab	ELISA test	positive
Lab	ELISA test	positive
Lab	Western blot	positive
Exam	Detailed physical	
Medication	AZT	

The idea of data segmentation is to provide patients a form of privacy by sequestering some data. Attempting to do so usually generates an inference problem, where the missing information can be learned from what remains. In health data, the inference problem is compounded because information is replicated in autonomous copies of information held in different organizations.

Shown above the HIV diagnosis is omitted, but could be inferred from the having two positive ELISA lab tests followed by a Western blot, or alternatively, from the AZT prescription. These other data elements are not limited to the provider’s EMR, but exist in autonomous data collections held by the lab, pharmacy, insurance companies, and others.

**Figure 7. Care and Safety (Example of an NHIN Trust Issue)**

An overall goal of the NHIN is for providers to use consolidated information over time, compiled from different providers to improve decision-making for the patient. Lack of trust by patients or providers will result in missing or incorrect data, and will thereby undermine the entire effort.

Here is an example.

A 24-year-old woman, Eve, sees Dr. Faye after obstetrician noted a murmur on prenatal examination in her 21st week of pregnancy. Denies any symptoms. Reports no history of allergies. Dr. Faye recommends an endocarditis prophylaxis and prescribes Biocef, orally. Life-threatening complications result because Eve did not remember, and Dr. Faye did not know, that Eve has a penicillin allergy with an immediate hypersensitivity reaction.

Our goal is for the NHIN to offer the kind of consolidated information that appears in (a) below. It shows that at an earlier time Eve had a prescription for penicillin and the next day presented at an emergency room.

10/27/2005 <sup>2</sup>	Pharmacy filled prescription for penicillin.
10/28/2005	Emergency room visit: diagnosis and procedures consistent with allergic reaction.
1/3/2009	Diagnosis of pregnancy
...	... visits related to pregnancy

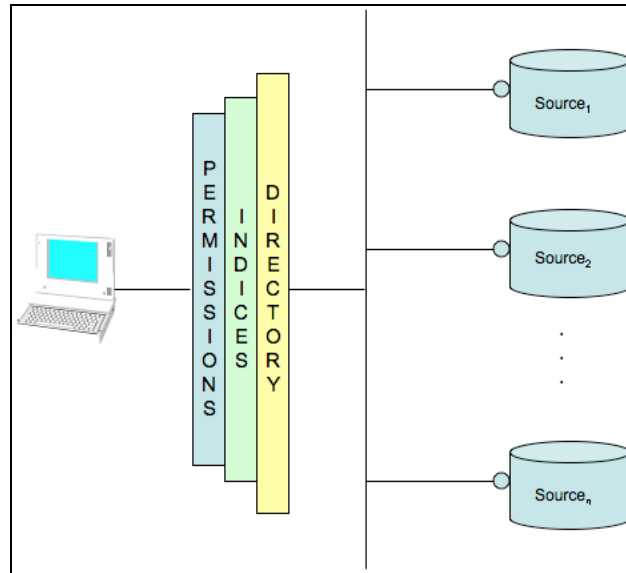
(a)

10/27/2005 <sup>2</sup>	Pharmacy filled prescription for penicillin.
1/3/2009	Diagnosis of pregnancy
...	... visits related to pregnancy

(b)

Of course, Eve could inform Dr. Faye that she never took the penicillin and presented in the emergency room for a different reason. If so, the data on its surface is misleading. It can also be misleading if information is missing. For example, consider the consolidated information in (b) above. If the emergency room visit happened but was not included in the consolidation, it gives Dr. Faye reason to believe the opposite, that Eve is not allergic to penicillin.

**Figure 8. Design Pattern 1: Global Query**



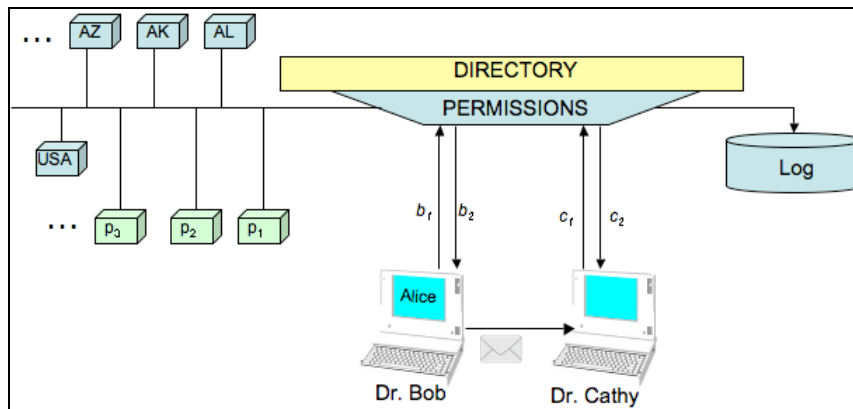
We introduce the term “Global Query” to describe an NHIN design capable of answering queries across a national collection of participating health information repositories in real-time, and providing the same answer as if all data were centralized. Indices may exist to speed operations.

The figure below depicts an instantiation of a Global Query design with global services (*Directory*, *Indices*, *Permissions*) connected to health information data sources (*Source<sub>i</sub>*). *Directory* interprets data addresses (similar to Domain Name Service on the Internet); *Indices* relate search criteria to data addresses; and, *Permissions* determine which results can be viewed based on the authentication and authorization of the requester, regulations, laws, and possibly patient consent. Global services could alternatively be de-centralized (see Figure 9). The small circle affixed to each data source represents some software necessary to operate at the site of the source.

In terms of use, web searching provides a quick analogy. In web searching (e.g. Google), a search string is entered and the search engine returns web addresses (URLs) of web pages that contain the search string. In Global Query, the results from a search string could be pointers to matching data contained in data sources. For example, searching “Alice Jones” might return pointers to all locations having information on Alice Jones if the recipient were permitted to know of such information.

Search is not the only kind of query supported. Other forms of queries may provide real-time numeric results, such as counts, regressions, and percentages –which of course, may have many public health uses. Another form of queries is data extraction, in which copies of information may be retrieved for secondary uses.

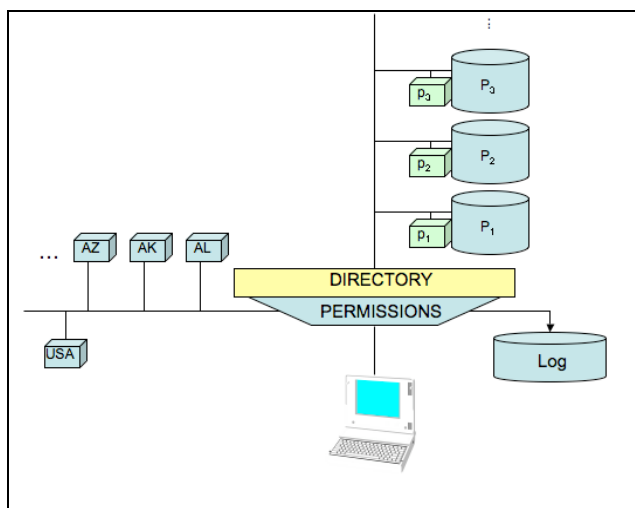
**Figure 9. Design Pattern 2: Certified Delivery**



We introduce the term “Certified Delivery” to describe an NHIN design capable of point-to-point delivery of health information with verifiable and accountable endorsements of sender and receiver per contents. Permission and logging services are usually essential in Certified Delivery designs.

The figure above depicts an instantiation of a Certified Delivery design with global services (*Directory*, *Permissions*, and *Log*) that enable the transfer of information about a patient (Alice) from one provider (Bob) to another (Cathy) if and only if the providers have adequate permissions to send and receive the information (i.e., Bob has permission to send Alice’s information to Cathy,  $b_1$  and  $b_2$ , and Cathy has permission to receive Alice’s information from Bob,  $c_1$  and  $c_2$ ). As in Figure 8, *Directory* interprets addresses (so that Bob’s reference to “Dr. Cathy” provides the address of her machine). Regulations, laws, and possibly patient consent, as well as the authentication and authorization of providers determine permission. In this instantiation, however, *Permission* is de-centralized (see Figure 8 for centralized version), allowing separate authorities to make decisions in accordance to the regulations in each state, federal regulations, patients by group ( $p_i$ ), and so on. A global *Log* service (which could also be de-centralized) records all requests and outcomes. Certified Delivery is analogous to having providers email and fax patient information to each other, but with more accountability and oversight.

**Figure 10. Design Pattern 3: Patient Central**

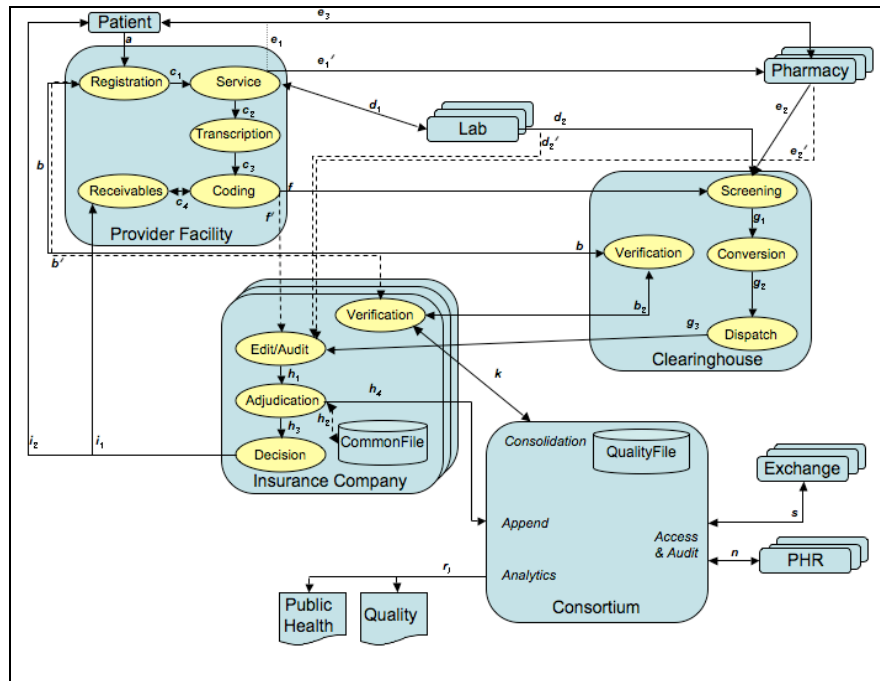


We introduce the term “Patient Central” to describe an NHIN design that uses globally available storage of patient information arranged by patient. The physical analogy is to one large file cabinet with files arranged by patient. Electronically, these would be globally accessible electronic medical records (EMRs), certified through the ONC approved certification process for EMRs, and therefore capable of storing clinical information in the same manner as the provider’s own EMR.

An EMR should not be confused with the general notion of a Patient Health Record, or PHR, in which a patient has a copy of his health information and may edit, annotate, or append it and the notion of patient ownership dominates. (Microsoft Healthvault and Google Health are PHRs). At least two companies, Kaiser Permanente and eHealthTrust, are EMRs fused with patient facing PHRs.

The figure above depicts an instantiation of a Patient Central design with global services (*Directory*, *Permissions*, and *Log*) and with de-centralized global storage ( $P_i$ ) that stores patient records with each patient having one and only one record across the set of repositories. A patient could have access to some or all of his information and could establish permissions for further sharing ( $p_i$ ). For example, patients could reveal medical information to researchers to determine possible eligibility for a clinical trial without revealing identity and researchers could search for clinical characteristics of interest, and could solicit participation based on matches, without knowing the identity of potential subjects.

**Figure 11. Design Pattern 4: Medical Billing Backbone**



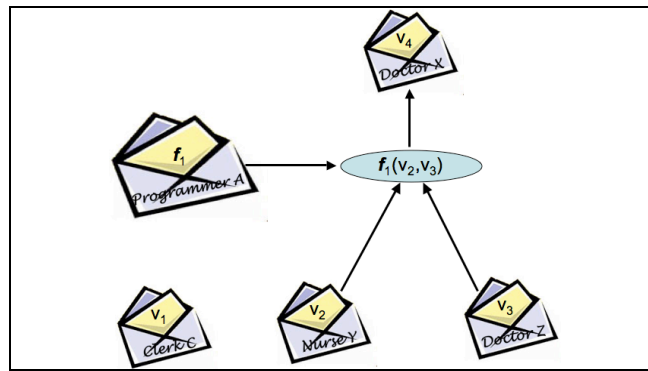
We introduce the term “Medical Billing Backbone” to describe an NHIN that uses the medical billing network for health information exchange. The existing medical billing framework has national connectivity, and its existing communication, data, and authentication standards may be expandable to serve as a necessary backbone able to carry lightweight information. Doing so could offer advantages in achieving some meaningful uses: (1) providers are already “wired” and using the billing framework, processing billions of claims a year; (2) the mechanism to check patient insurance eligibility may expand to provide relevant patient information (e.g., problems, medications, and allergies) at the time and place of service; (3) CMS has a national program (Physician Quality Reporting Initiative) that already captures quality measures through claims processing; and, (4) payment incentives on claims can drive ongoing provider participation. Emdeon and others have posed examples.

The figure above depicts the workflow from a single patient-provider encounter through the medical billing framework and adds a Consortium to orchestrate operations and sustain interoperability (the latter being similar to W3C’s oversight of the Web). Principal entities appear as rounded rectangles, business functions as ovals, and information flows as edges.

Insurance eligibility check when patient arrives for service ( $b-b_2-k-b'$ ) provides clinical information to provider. Adjudicated claims store lightweight clinical information ( $f-h, f-g-h, d-h, d-g-h, e-h, e-g-h$ ) from provider.



**Figure 12. Design Pattern 5: Heavy Data**



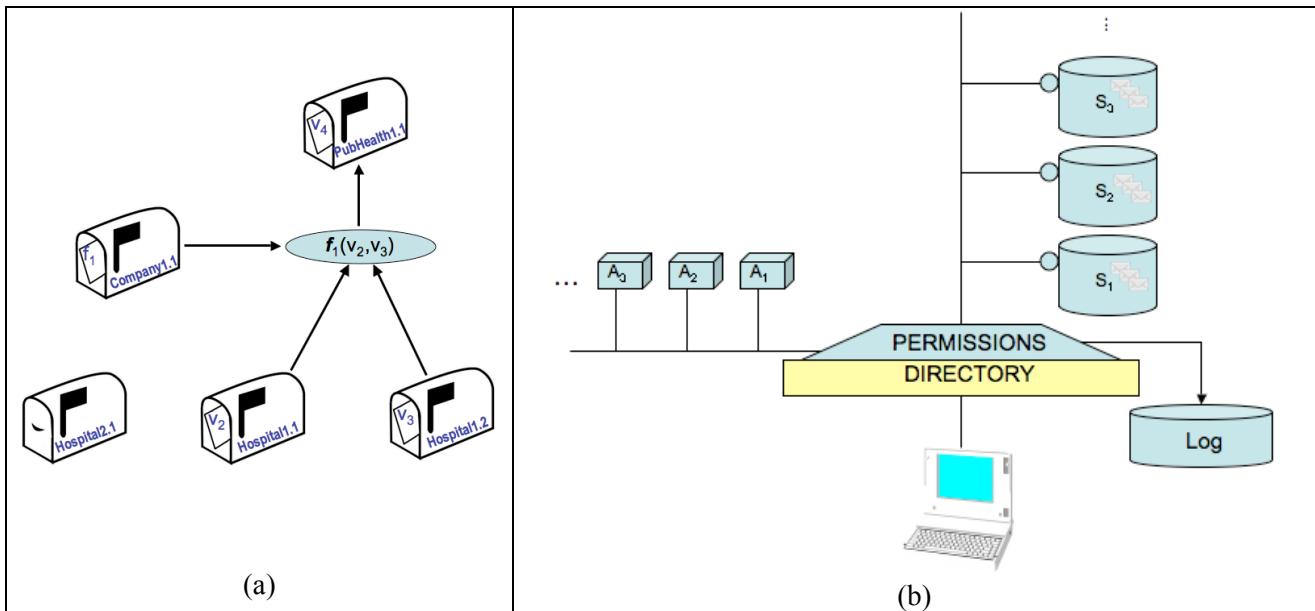
We introduce the term “Heavy Data” to describe an NHIN design that tethers or embeds meta-information along with data values. During information exchange, meta-information would travel with its data, and depending on the kind of meta-information used, may even be changed as a result of data sharing. Examples of meta-information include digital signatures and access policies.

There are several kinds of possible meta-information to consider, which should be used, if any, depends on what is to be optimized or accomplished. For example, we introduce the term “dependency tracking model” as a possible way to add retrospective accountability to data sharing by having all data values, functions on data values, and resulting values cryptographically signed, thereby affixing a certifiable reference to each. If there is a subsequent challenge to the validity of a value, even one resulting from a computation, the accompanying certificates should provide an accounting trail of entities responsible for the resulting value. The figure above offers a depiction of dependency tracking. Each component (value, function, or result) is signed by the entity responsible for collecting, composing, or computing it. The value  $v_2$  is signed by *Nurse Y*, and  $v_3$  by *Doctor Z*. *Doctor X* combines these values using function  $f_1$  to produce  $v_4$ . If the validity of  $v_4$  is suspect, the digital signatures of the components on which it is based ( $f_1$ ,  $v_2$  and  $v_3$ ) provide an audit trail.

Another example of meta-information is “sticky policies”, which are policies that may be affixed to data values to possibly represent data sharing allowances, prohibitions and consents. Heavy data can of course be combined with other design patterns described herein (e.g., Design patterns 1, 2, 3, and 6).

The figure above provides an instantiation of one form of Heavy Data –dependency tracking using digital signatures globally, thereby seeming to enable each value to report its own audit trail and identify its originating source.

**Figure 13. Design Pattern 6: Pointer Addressing**



Pointer addressing (a) and its use with heavy data using grid computing technologies (b).

We use the term “Pointer Addressing” to describe an NHIN design that shares published network addresses of data rather than data values themselves. Because meta-information tends to be costly to move and because error correction in data replication tends to be difficult or impossible, one idea is to keep the data close to collection. The source maintains the value at a published address. Others who have permission to access the value can do so and work with cached copies, but if the value is corrected, subsequent requests use the updated value and the audit log could identify all who could be notified.

The figure (a) above shows a depiction of pointer addressing using initial values ( $v_2, v_3$ ) published by a hospital at addresses *Hospital1.1* and *Hospital1.2*, a function ( $f_1$ ) published by a vendor, and a resulting value  $v_4 = f_1(v_2, v_3)$ , published by a public health department at *PubHealth1.1*. Changing a hospital value updates the public health computation. *Hospital2.1* has a value that is not accessible.

The figure (b) above shows a Global Query design using heavy data (access policies attached) and pointer addressing. Adjudication of data’s policies with requester’s credentials by a certified adjudicator ( $A_i$ , in shown decentralized case or could be centralized) determine access decisions. The National Coalition for Health Integration, UCLA Health System and St. Johns Health Center have a version of this in operation using FTP-like service over grid computing.