

The National Coordinator for Health Information Technology
Privacy and Security Tiger Team
Health Information Technology Policy Committee
Patient Linking Hearing
December 9, 2010
Washington, D.C.
CIGNA
Paul Oates

1. BACKGROUND – WHY ARE HEALTH PLANS IMPACTED BY THE IDENTITY ISSUE TOO?

CIGNA is a global health services company that offers group and individual health insurance, health management services such as wellness programs, coaching and disease management, disability and other types of health related services. In addition to traditional health plan services, CIGNA now offers onsite health centers, pharmacy services, one-on-one health coaching, health promotion, health improvement programs, educational seminars, biometric screenings and flu shots as we help consumers improve their health and well-being.

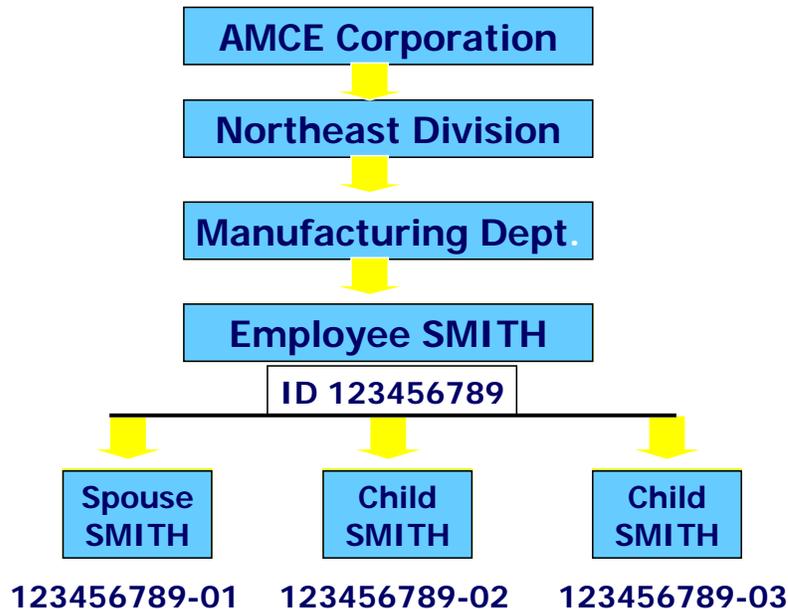
Healthcare reform impacts health plans' fundamental business models, changing from a business-to-business model (where employers negotiate and supply coverage) to now include consumer-direct purchase of services. Additionally, health plans are now infomediaries that offer care coordination, informatics, and other value added services to improve quality and lower cost of care. As CIGNA focuses on health services, CIGNA (and other health plans) are focusing much more on accurately tracking interactions with consumers. This requires strong identity solutions. But this also has challenges in most health plan data architectures, outlined below.....

Historically at most health plans, the identity of an individual has meaning within an employer. This is because health plans like CIGNA historically have focused on the employer as the customer unit of work, not the individual person. Because of this, the business (and information) architecture of most health plans was tied to the employer and the person identifier was a sub-component of employer.

The historical purpose of keeping person data was to derive eligibility for benefits and pay claims, not primarily to improve care or service an individual. So, the primary data attributes linked to a person largely revolved around tying a person to their dates of eligibility, their plan type and features.

The basic data about a person is also supplied by the employer, not the individual, usually using payroll data. Sometimes the data is not complete or current. For example, some employers do not collect home phone number or have sketchy details about all of the members of a family.

The basic business view of a person in many health plans, therefore, looked like this:



In this typical example, an employer, shown here as Acme Corp, has several divisions and sub-components. An employee – Smith - is attached to one of those subcomponents because the employee (or subscriber) elects coverage. A spouse and children are subcomponents of the employee – they do not stand on their own.

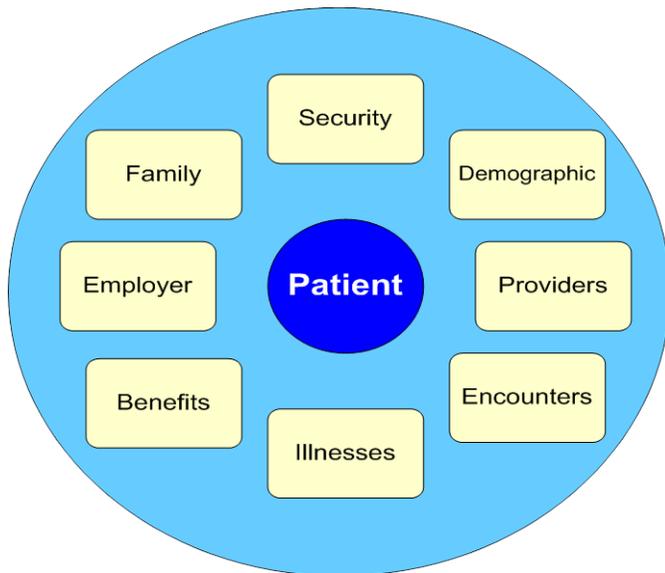
This structure causes identity challenges, depending on the type of system managing this data. For example, if a person ceases to be eligible for a short time, their record could become inactive – and their dependents too - even though from a servicing or care point of view they still may be within an episode of care or interaction. Another example: Spouses sometimes work for another employer covered by the same company, and therefore could appear twice. Subscribers may also work for more than one company or have government/Medicare benefits, and therefore appear multiple times, but with different identifiers. All of this is intentional, as coverage ties to the instance of a person’s tie to an employer and their coverage record. This structure makes identifying who *person* Smith is challenging, as all of these instances of Smith can have different identifiers.

1.1. Future State.....Today

The traditional view of health plan and health care professional / hospital system roles is changing.

- Healthcare Professionals and Hospitals (aka Providers)
 - Primary role, historically: Diagnosis & Determine Course of Treatment
 - Additional role: Holistic patient management, necessitating linkage of data from across stakeholders
- Health plans
 - Primary role, historically: Transaction manager...eligibility, claims
 - Additional role: Information provider, care coordinator, necessitating linkage of information from across stakeholders

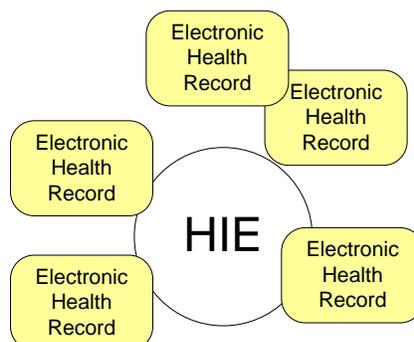
As a result of this change, many health plans are increasingly focused on building a view of a person that links all of their various attributes together in order to offer value-added information back to the provider and consumer. This kind of linkage results in a very different view of a person being created than the picture that was shown above, where the person's identity is at the center and characteristics like coverage are attributes of a person:



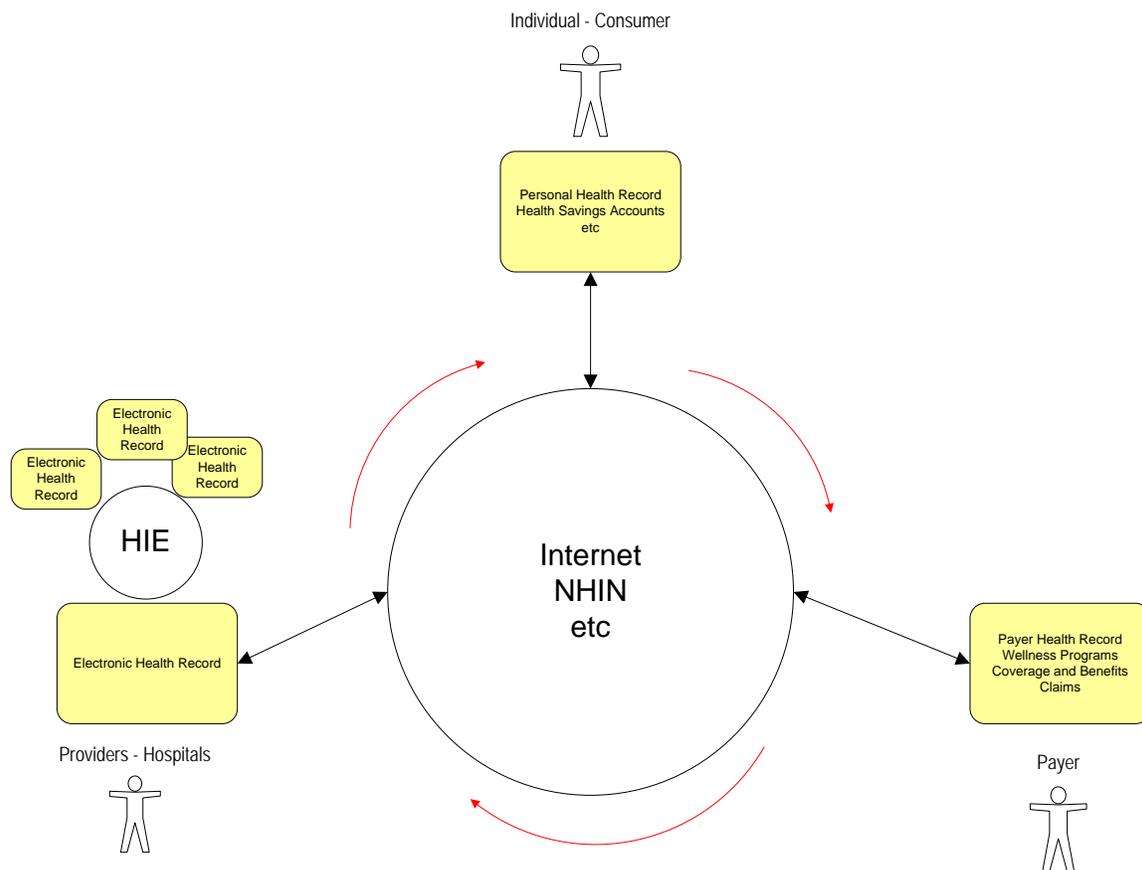
Where this leads us is to a realization that the identity linkage problem is not just within and between providers, but within the health plan too.....and also whenever information moves between health plan, provider and consumer stakeholders. This requires a slightly different view of "what's the problem."

2. A TWIST ON "WHAT IS THE PROBLEM"

The Patient Linking hearing is largely focused on just that: patient linking. Meaning, linking of data about people who are flowing through provider and hospital systems – and an HIE – *as they are treated for illnesses*. However, this restricts the linkage discussion to a fairly limited view of the healthcare system:



The importance of linkage accuracy in this context cannot be understated and the impacts and risks of inaccuracy have been well documented. (see endnotes (1)). *However, the linkage problem occurs not only within the provider-hospital-HIE space, but also when information about a person crosses over to systems and information services managed by the health plan or other 3rd parties that contain person identity:*



Rather than “patient identity linkage”, perhaps a better context would be “consumer identity linkage”. When the individual is a patient, then linking identity is critical to patient care accuracy. But when the individual is playing the role of consumer trying to actively manage their care, or when information is flowing between the consumer, provider and health plan in, say, an Accountable Care Model, then the identity problem extends to all of the other places where identity flows. Our industry will not be able to successfully implement Accountable Care Organizations, Medical Home, or other collaborative care models unless person identity is linked across the entire ecosystem. A few examples:

- To meet meaningful use, a provider assesses a patient for one of the stage 1 core measures (stress, weight, tobacco use). Upon discovering a quality gap (for

example, patient exhibits high BMI), the provider requests information from the health plan regarding availability of a healthy eating program (using the subscriber's number). The BMI (linked to the provider's patient identifier) is added to the person's health plan health record (keyed by the health plan's identifier for the person) and the weight program is started should that person agree to the program by adding it to their personalized care portal at the health plan (linked to the consumer's log in and identifier).

- Within a health plan data warehouse, a claim (containing a subscriber number) is linked to biometric information (identified with a medical record number) and lab results (identified with a different patient identifier from the lab). The resulting longitudinal view is used to produce actionable gaps in care information to assist in care coordination.
- A provider requests a CCD from a health plan (using the patient's subscriber ID on the ID card). The health plan returns a CCD that contains information that was integrated from multiple historical health record documents (which came from other sources that have various identifiers). Some information may have come from the patient's own PHR if consent was obtained (which is keyed by another identifier perhaps). This is all linked and returned to the provider. The provider then links the information back to the patient identifier in their own health record.
- Multiple health plans that have historical information about a consumer – all keyed with different identifiers – return information to an HIE to establish a centralized record set, which is then keyed by a new patient identifier for the HIE. Later, during an illness, a care coordinator from one of the health plans and a care manager in the hospital system exchange information which is linked together at the time of the transaction to bridge between provider and health plan.
- A consumer obtains their health information from the provider which is keyed by their patient number to add to their own, unteathered, PHR. The consumer also logs onto the health plan portal (via another identifier) and requests a download of their claims and health savings account transactions in order to assess the cost of their care and options for future care. The identifiers on the provider and health plan records are different, requiring the consumer to keep track of them and correctly request their information from each party.

The examples are negotiable and could cause robust discussion at a future session. What is more certain is that linking clinical information across health plans and providers (and others) will increase. Providers, hospitals and the HIEs have data that is very deep clinically, but perhaps not broad. Health plans have data that is very broad, but sometimes not deep. Only by assembling the two views is a composite of the cost and quality of care achievable. And consumers will increasingly be required to navigate between components of the health ecosystem in order to manage their own care. Or a health services company like CIGNA assists consumers with navigation. Each flow between stakeholders (provider, health plan, PHR vendor, disease management company, etc) requires identity. Consumer identity linkage is the lynchpin.

3. WHAT ARE THE SOLUTIONS?

To enable CIGNA to accurately link information and accurately interact with consumers and providers in new care delivery models, we are investing in people, process and technology solutions to increase accuracy and quality of identity. While we can't speak for all other health plans, it is knowable that many other health plans are making similar investments. These solutions and approaches include:

1. Installation and use of advanced probabilistic matching software. This type of software has been described by others, so no additional explanation needed here (see endnotes). What we can say is that these tools scale very well and have a high level of accuracy. We have tens of millions of records loaded into a matching tool and matching performance is highly responsive and accurate, once tuned.
2. Data stewardship people, process and technology. This includes establishing staff roles who govern data definition and use, including the data elements used in person matching at an enterprise level and ensuring that identifiers and identity-related data elements adhere to corporate and industry standards.
3. Common data dictionaries and canonical models. This implies development and use of common data definitions in all applications related to patient identity and consistency of interface definitions when identity moves between systems
4. Tools and people to measure data quality at rest and when data moves. There are tools that enable periodic measurement of data quality aspects such as completion of fields, correct formats, and frequency/distribution analysis. Findings are cycled back into operational systems improvements over time. The same toolset can be used for monitoring quality on an ongoing basis and issuing alerts so that corrupt identity data can be caught before use.
5. Measurements of matching algorithm accuracy / duplicate rates by source system. This is challenging, as it is difficult to measure when the right thing did not occur (false negatives).
6. Overall governance. Ensuring the identity has an owner accountable for creation, updates, and read activities. We have a group of people whose job it is to handle unclear matches and clean up duplicate data.

Most of these solutions are typical of *master data management techniques* applied in specific ways related to identity management.

3.1. Status of Solutions

These solutions are in various states of maturity not only in CIGNA, but likely in most health plan and non-health plan organizations.

3.1.1. What are standards for identifying individuals?

We require a minimum of first name, last name, date of birth, and gender in order to create a person record or run a match algorithm in a Master Person Index (MPI). This is a minimum. Address and source system identifiers like member number, subscriber number, SSN obviously improve the match. Note carefully that we are – like many others – transitioning away from SSN as an identifier. It's required we do so and also the

right thing to do. However, a replacement identifier that offers similar value has not yet been made available (see universal patient ID comments below).

We currently see de-duplicate rates across source systems from 6% to 35%. This rate is solely based on the specific characteristics of a source system's processes and data quality. Fallout task creation – where a match is not clear – occurs at low levels once algorithms are tuned (much less than 1%).

3.1.2. How can we be sure we are accurately linking?

The key to accuracy is testing and algorithm tuning. Before a source of data is installed, algorithms are tuned for a source of data to accommodate any data accuracy issues. Each match is inspected by a human and verified. Once the algorithm for that source is installed, we put a monitoring process in place. And then we use the same algorithm and tool for real time matching at point of use.

3.1.3. How do you resolve ambiguous matches?

A data stewardship team is necessary to clean up ambiguous matches.

3.1.4. What is done about no match although one was expected? (false negatives)

Testing and inspection validates algorithm performance. Once an algorithm is in production, it is trusted. However, the ability to measure what didn't happen – false negatives in other words – is an issue with no clear solution. See gaps list below.

3.1.5. What's the consequence of a wrong match?

The consequences of a wrong match are well documented (see (1)). Additional research could be pursued to determine how often these consequences actually occur. In many cases, the results of a match are used by a human who is accountable for applying the information to the context of the situation they are in. If information does not look correct, it is possible that users of that information will correct it if there is a process available for doing so.

3.1.6. What should ONC do to address patient matching problems?

1. Establish a set of best practices – see below
2. Train and inject those best practices into the industry
3. Extend the range of the current discussion to include health plans and other stakeholders who are part of the patient care coordination process. This implies that:
 - a. The HIT Policy and Standards Committees would add ACO, Medical home and other care coordination processes and profiles to standards development that include clinical data exchange between providers, health plans, pharmacy, and others. Today, standards development is essentially 100% focused on patient illness management. For example, we need IHE profiles that include receipt of gaps information from a

health plan into the electronic health record, along with the associated identity requirements that would entail. Another example: profiles that link stage 1 core meaningful use measures (stress, weight, tobacco use) found within an individual to that individual's wellness programs available at a health plan.

- b. Extending the NHIN discussion to include health plans and others who are not solely in the patient care delivery space, but who will become nodes on the NHIN in order to effectively administer Accountable Care Organization or Medical Home models
 - c. Establishing info-mediaries as new stakeholder types within the standards discussions. ACOs, Medical Homes and quality measurement/comparative effectiveness research requires more study to establish best practices for identity management across stakeholders.
4. Encourage consistency. Use the bully pulpit to make identity management investments as important as the health record investment
 5. Pursue a universal identifier as a replacement for the role previously played by SSN in linkage. But acknowledge that it is only a helper, not a total answer to the problem. The Rand Study suggested this as a transition, but it's likely that a hybrid approach of consistent identifiers and probabilistic matching will be the solution long term.

3.2. What Are the Gaps?

In general, there are several gaps and challenges that will require coping strategies:

1. Health plans will need to maintain two related identities of a person for the foreseeable future (a "dual view"): a) an identity that describes a person as part of an employer who has purchased a health plan b) an identity of the person that ties together all of their employer, provider and other consumer relationships into one. For example, returning eligibility on a HIPAA 271 transaction requires use of an identifier tied to a person's coverage. But returning a CCD that contains all of the relevant clinical information about a person to a provider requires use of an identifier that links multiple instances of a person's history across employers and all clinical settings. Eventually this dual-view may converge.
 - a. Interesting question: The vision for health exchanges includes the ability to allow consumers an open market for health coverage. Therefore, person identity will be tied to benefits and coverage. This suggests that health exchanges will need to cope with the same identity issue that health plans cope with today if we intend to build bridges between the insurance exchanges and the HIEs that also exchange information about consumer.
2. The industry will always have the need for matching, even if we someday have a universal person identifier. Like we have seen with NPI, the use of the identifier can be interpreted differently, used in different contexts, and, sometimes, mistyped. It is unlikely that a supplier of information would simply take an identifier from a requestor and do a lookup in their own systems using that identifier without applying some level of matching to ensure that the identifier was an accurate key to the requested information. And in the specific case of a

- provider requesting clinical information from a health plan who is doing health management, we'd always feel more comfortable running that request through a probabilistic algorithm before responding. Pix/PDQ has a long life expectancy.
3. Most of the attention is given to person identity, but how best to link households together also will need attention. Often, care decisions apply to those who live with the individual that has an illness. Or, care programs may be offered not only to a primary person in a household, but to all of the other people related to that person. How to define and create identity for a grouping of individuals requires more study.
 4. Allowing consumers to choose their own identifiers and/or enabling the employer to use the individual's workplace identifier compound the problem of identity linking. This compounds the problem of accurate identity because of duplicates (the same employer generated identifier might be used by multiple employers). We don't know where consumer empowerment will lead and whether consumers will dictate what their identifier will be.....we've seen the challenges of lack of consumer standards for email addresses. It's rather daunting to think the same thing could happen if consumers govern their identification choices too.
 5. Probabilistic matching tools still have some work to do to provide statistics on matching performance. It is not really practical to keep audit records for all matches and non-matches due to the granularity of decisions and the scale of these systems. However, more research into how to leave a trail of breadcrumbs – especially for false negatives – in order to measure accuracy is needed. It's hard to measure what didn't happen.
 6. HIE learnings are scarce. Most information accesses to health plans are still via HIPAA transactions that use the member number on an ID card. We do not have significant experience yet on how person identity on clinical information requests will flow from a health record system, across the HIE, to the health plan, and back. CIGNA is running pilots to learn about this and would expect to gain experience as health record system adoption increases.
 7. Another requirement for accuracy is that each source system from which we acquire information is required to provide a consistent identifier for a person entity. This is inherently difficult, because it implies that source systems like labs, eligibility systems, pharmacy systems themselves have robust matching capabilities when in fact many are deterministic.
 8. There is not a full-scale model of identity flows across the entire health ecosystem. A taskforce should take on the task of building a data flow model across the health ecosystem so that the industry can quantify the number of identity linkage occurrences within key ACO and Medical Home processes.

4. POSSIBLE BEST PRACTICES

It is unlikely that definitive standards for patient matching can be established across the industry. There is too much variability in data quality and processes in organizations that would prevent a one size fits all standard. But it would be helpful to establish a set of best practices and then inject those practices into the training and education processes that are under development in places like regional extension centers and add those best practices as criteria for certification in some cases. Suggestions for best practices:

- Use advanced probabilistic matching software. Discourage use of deterministic matching tools and processes. This not easy and can be expensive, but necessary.
 - o Perhaps open source probabilistic matching capabilities can be made available or commercially available tools can be more commonly baked into transactional systems
- When implementing, perform many trial runs of identity linkages to tune algorithms based on populations and data quality represented in source systems. Use humans to inspect results and cycle results back into tuning decisions. This means health plans, HIEs, and others need to hire and train data stewards. Perhaps data stewardship and identity expertise should be part of ONC sponsored training.
- Using probabilistic matching tools is not enough. Organizations also need to treat data as a significant asset. This means creating and adopting data stewardship, data governance, data quality measurement methodologies and applying tools and techniques to manage data that impacts person identity. This means hiring information architects and driving info analysis tool use.
- Require that source systems keep identity consistent for a person over time. If John Smith is ID 123 today, John Smith needs to be ID 123 tomorrow. If this is not accomplished, then John Smith will look like a different person each time his data is exchanged. Accomplishing this would require some form of “standard” to test health record systems to verify that identity can be managed consistently for an individual and that controls are in place to ensure that a person’s identity cannot be inadvertently changed by a user. Consistent approaches for linking and unlinking would be included in this topic. See next item.
- Make sure MPI systems truly link and unlink records, not merge. The ability to keep track of history and the components of a person’s identity needs to be maintained and the ability to undo a linkage must be available. This can’t happen if merges occur in MPIs. Merges only should occur in transactional sources of record.
- Ensure that the algorithms – if not the actual tool – used to populate a master person index is exactly the same algorithm used when reading that index. It does little good to probabilistically create an MPI only to allow deterministic systems and processes to read it.

5. CREDITS AND NOTES

- (1) Patient Integrity White Paper. HIMSS Patient Identity Integrity Workgroup. December 2009 and the RAND Health Study – Identity Crises – Costs and Benefits of Unique Patient Identifier.