Accurate and Trusted Data-
The Foundation for EHR Programs

Susan J Hyatt
President and CEO
HYATTDIO, Inc.

Lorraine Fernandes, RHIA
Global Healthcare Ambassador
IBM Information Management

Healthcare priorities and goals

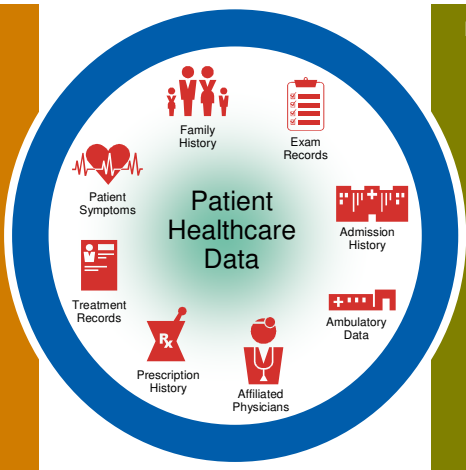
- PCEHR goal: "Improve accessibility to health services and patient information significantly enhancing health outcomes"
- HI goal: "Give individuals and healthcare providers confidence that the right health information is associated with the right individual at the point of care"

Laudable goals that require a strong technical, governance, and stewardship platform

Goals and Registries: Highly complimentary

Patient Registry

- Locate and link same patient records across systems
- Improve strength and reliability of health identifiers
- Search & retrieve patient identity in sub-seconds
- Improve data quality with tools and auto reconciliation
- Support patient privacy and access policies



Provider Registry

- Create a single view of providers and their organizations
- Facilitate consistent and timely provider communication
- Support privacy by enabling correct results routing
- Operationalise data stewardship to improve data quality
- Identify patient and provider relationships for case management

HI Service and Deterministic Matching: Challenges

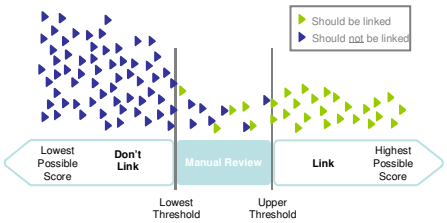
- Exact match based upon byte to byte comparison of data
- Rules approach actually counter-productive
  - No match, no return,
  - Multiple candidates, no return
  - Can't address variability in data given thousands of users and sources
- Not tunable
  - Can't establish minimum confidence level or threshold for a match
- No consideration of data quality
- Very high missed match (false negative) rate
- Can't scale as volume, systems, environment changes

Probabilistic Matching and Searching

- Based upon mathematical principles of likelihood ratio that considers the quality, uniqueness & frequency of the data
  - Offsets data transposition, typographical errors, homophones
  - Uses all data available to define a potential match
  - Efficient, as single search based upon input of available data
- Ranks candidates based upon likelihood of a match—it is *NOT* phishing
- Configurable by customer based upon business environment, goals, and objectives
  - Single threshold presenting only BEST match should be used by Medicare Australia
- Probabilistic mimics human brain in resolving errors

Summary of Matching Steps

- **Step 1**: Optimizes data for statistical comparisons
  - Normalizes & compacts data, creates derived data layer, source data remains intact
  - Phonetic equivalences, tokenization, nicknames, etc.
- **Step 2:** Finds all the potential matches
  - Casts a wide net – all matches on current or historical attributes, prevents misses
  - Partial matches, reversals, anonymous values, etc.
- **Step 3:** Scores accurately via probabilistic statistics
  - Compares attributes one-by-one and produces a weighted score (likelihood ratio)
  - Frequency weights specific to your business
  - Edit distance, proximity of match
  - Allows custom deterministic rules
- **Step 4:** Custom threshold settings
  - Single or dual threshold models
  - Link, don't link, don't know – "learns" from manual input
  - Manage cost/quality trade-offs
  - Manage the linkages, workflow review



Search Results

Search Results - Entity — Displaying items 1 to 6 of 6

Inspect	Score	Tasks	EID	Patient Name	Home Address	Home Phone	SSN	Date of Birth	Gender
	6.9		10271	POTTER, BOB, ALLEN	12 WINFIELD, San Francisco, CA, 90313	310-967-5287		1960-07-09	M
	6.8		1639	POTTER, ROBERT,	12 WHINFIELD PL, San Francisco, CA, 90313	310-976-2587	952169663	1960-07-09	M
	6.2	---	223	MOLINERO, ROBERTA,	39 ALBERT RD, Los Angles, CA, 95220	209-242-9413	933181631	1960-07-09	F
	1.9	---	5519	PEDDER, ROBERT, HAROLD	134 DUNKIRK LANE, Chatham, CA, 97025	213-486-4448	928338234	1936-10-12	M
	1.3	---	4859	ROPER, ROBERT, PETER	50 WOODPLUMPTON RD, Hearst, CA, 90017	323-274-9181	929574975	1960-07-09	M
	1.3	---	4391	BULLOUGH, PETER, ROBERT	16 GORDON ST, San Diego, CA, 92032	213-519-3833	903391979	1961-09-25	M

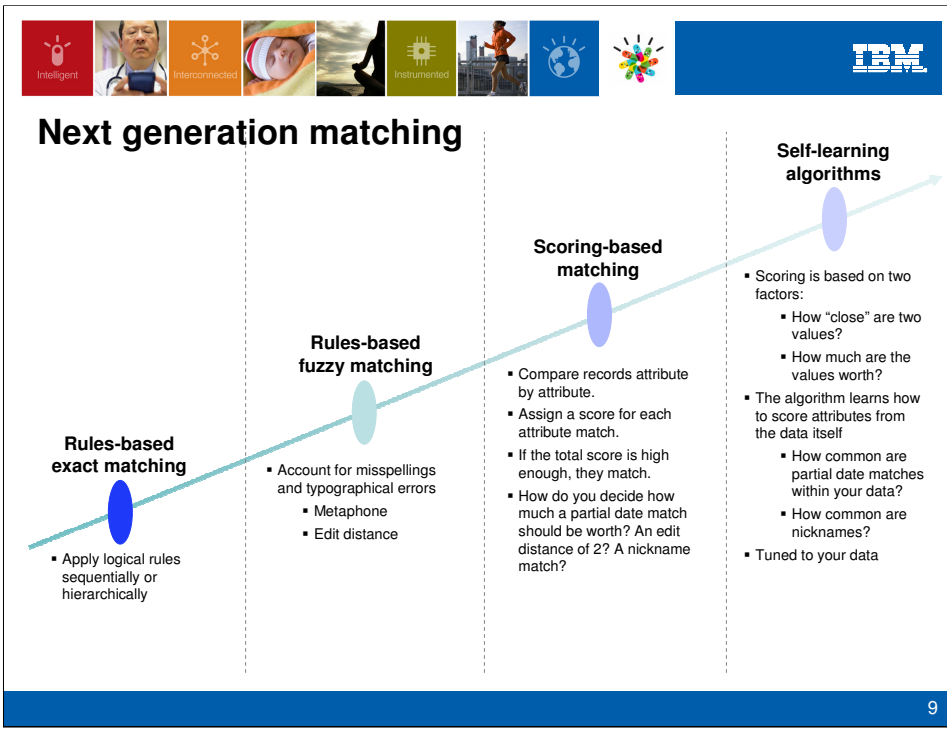
Fundamental Approach differences

Deterministic

- No sensitivity to data quality
- Best with small databases of less than 1M records
- Requires significant IT support to write extensive rules, and results are limited
- Binary, only answer is yes or no

Probabilistic

- Considers data content and data quality in matching
- Scales to billions of records
- Configurable to business goals, environment
- Technology does the “work”


Intelligent
Interconnected
Instrumented
IBM
Next generation matching
Rules-based exact matching
Apply logical rules sequentially or hierarchically
Rules-based fuzzy matching
Account for misspellings and typographical errors
Metaphone
Edit distance
Scoring-based matching
Compare records attribute by attribute.
Assign a score for each attribute match.
If the total score is high enough, they match.
How do you decide how much a partial date match should be worth? An edit distance of 2? A nickname match?
Self-learning algorithms
Scoring is based on two factors:
How "close" are two values?
How much are the values worth?
The algorithm learns how to score attributes from the data itself
How common are partial date matches within your data?
How common are nicknames?
Tuned to your data
9

Best Practices for Health Identifiers

- Use probabilistic search to augment HI service.
  - Accommodates data quality issues, user expectations, *and supports privacy requirements.*
  - Supports multiple use cases through configurability
- Design for today, but keep your eyes on tomorrow
  - Scalability, multiple use cases, HC transformation
- Design a system that supports auditing, analytics
- Build processes at the onset for identifying, reporting and correcting errors.
- Processes and technology should support validating data

No system is perfect, so build plan with policy, processes and technology to correct

NHS Wales: Informing Healthcare Programme

Challenge

- Create a central registry to enable accurate patient identification in support of national eHealth infrastructure
- Support a federated architecture for participating Local Health Boards (LHB) and providers
- Improve data quality in existing Welch Demographic System (WDS)

Solution

- IBM® Initiate® Patient will be the national master patient index to facilitate information sharing across multiple health services
  - Created a shared MPI on the national infrastructure
  - Each LHB adding sources to their partition
  - Integrating data from patient administration, lab, radiology
  - WDS data Loaded as national baseline

Benefits

- Reduce creation of duplicate records by integrating with Patient Administration Systems (PAS)
- Create a unified view of the patient across patient administration, pathology, radiology, theatre systems & specialty systems
- Improve data quality across systems by identifying incorrect or incomplete records and recommending corrections to participating sources
- Enable the clinical portal by providing information about location of patient records that exist within various systems

Resources

- MDM and Accurate Matching
  - http://public.dhe.ibm.com/common/ssi/ecm/en/imw14331usen/IMW14331USEN.PDF
- Probabilistic vs. Deterministic Matching
  - http://www.information-management.com/specialreports/20070118/1071712-1.html
- Accurate Patient Matching, an Asia Pacific Perspective
  - http://public.dhe.ibm.com/common/ssi/ecm/en/imw14575usen/IMW14575USEN.PDF

CASE STUDIES

Case Study: Large APAC Government Health Agency

Challenge

- Highly controlled national ID for all citizens also used for health identifier
- Significant patient data duplication due to:
  - Errors in data capture during admission
  - Patients moving across clusters for treatment
  - Use of healthcare services by non-citizens
- Patient data could not be shared across different applications and healthcare delivery clusters as they used different internal identifiers
- Healthcare transformation underway required accurate, robust and immediate patient & provider identification to support a new national EHR

Solution

- Information Agenda for Healthcare
- IBM Initiate Patient

Benefits

- Enabled a comprehensive national EHR across the entire continuum of care including public and private sectors
- Improved patient care through the integration of patient records, realizing the government's vision of One Patient – One Medical Record
- Helped decrease costs through improved operational efficiency and regulatory compliance
- Improved patient experience by providing clinicians with critical information at the point of care

Case Study: Public Hospital Group in Belgium

Challenge

- Large hospital network with disparate systems
- 40% of patients treated in more than one facility
- No global view for admissions resulted in duplicate rates between 4 – 15%. At one facility, 35-45 duplicate records created daily
- New medical portal being put into place
- No health identifier on record for > 20% of patients
- Existing deterministic matching not sufficient

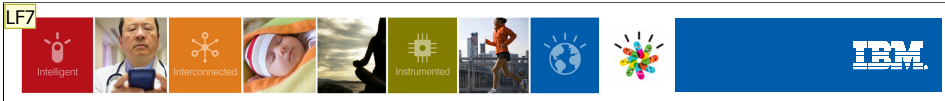
Solution

- IBM® Initiate® Patient uses probabilistic algorithms to match and link records across systems even when data is incomplete
- IBM® Initiate® Inspector creates daily task lists for data stewards to merge, link or correct duplicate records
- Real time search & retrieval will be implemented Q3 2011

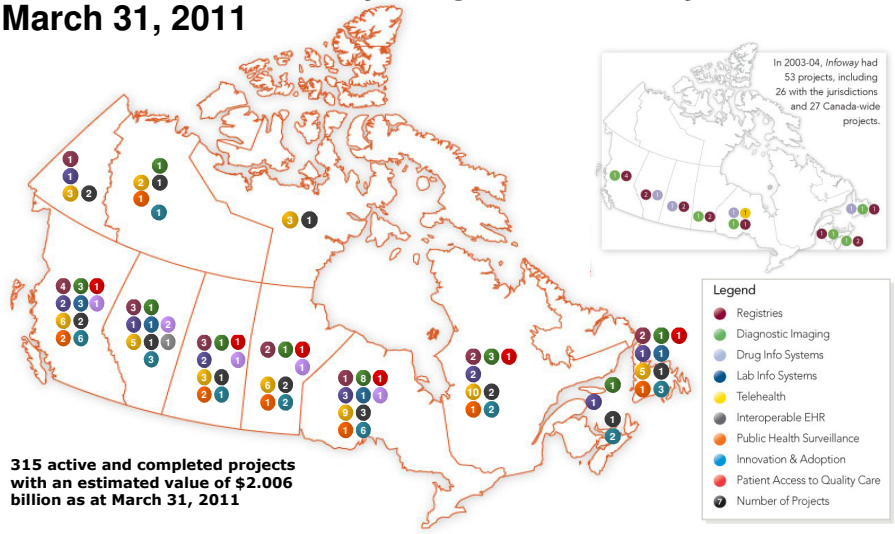
Benefits

- Improve and maintain high data quality
- Streamline process to identify, track and resolve data quality issues
- Create virtual patient record in real-time for clinical portal
- Enable timely and accurate sharing of patient data

LF7



Canada Health Infoway Program Summary: March 31, 2011



315 active and completed projects with an estimated value of $2.006 billion as at March 31, 2011

16

Slide 16

LF7 need za footnote that credits this ti Canada Health Infoway 2011
LF, 12/09/2011