Archiving for today...and the next century

*Our clinical information: When our present is past, what can the future learn?*

Mark Frisse, MD, MS, MBA
Vanderbilt University
April 5, 2011
Our charge

How to preserve and provide access of digital health information for a sufficiently long period of time to maximize its value to:

• patients
• caretakers
• scientists
• <others?>

Expected workshop outcomes

• Understand the current landscape on EHRs
• Survey current practices and identify best strategies to be used as models
• Begin to develop requirements, technologies, standards and best practices for long-term preservation and life-cycle management on EHRs
• Differentiate between requirements for patient care and those for secondary use
• Identify cultural and technological challenges
• Catalog current legal requirements for retention of EHRs
• Identify interested collaborators to form a WG on this area
• Discuss possible test scenarios and datasets for collaboration and test bed
What is the problem......today?
Digital preservation contexts

Our goal: to address the sustainability challenges & develop a course of action

- scholarly discourse
- research data
- commercially-owned cultural content
- collectively-produced health care content

Top four ways of losing digital information

- Can not find it
- Can not validate its authenticity
- Can not read it
- Can not interpret it correctly

Implications to care – today

• Can not find it
  – Clinical care (Hogan & Wagner, 1997)
  – Public health (Doyle, 2002)

• Can not validate its authenticity (see: authentication)
• Can not read it (see: interoperability)
• Can not interpret it correctly (see: medical errors)
Principles: *Ensuring Long-Term Access to Digital Information* *

• Recognition of the benefits of preservation
• Incentives to act in the public interest
• Selection of materials with long-term value
• Organization/governance of preservation
• Ongoing, efficient resource allocation

BRTF Principles – and clinical care

• Recognition of the benefits of preservation
• Incentives to act in the public interest
• Selection of materials with long-term value
• Organization/governance of preservation
• Ongoing, efficient resource allocation

How can we characterize the demand for information?
Time value of information (static media)
The general case (slowly dynamic media)

Value

cost / unit time

$ 

cumulative cost

$ 

time →
Time-value of information (active dynamic media)

value

cumulative cost

Value

$\text{cost / unit time}$

time $\rightarrow$

$\text{Time - value of information (active dynamic media)}$
Value of a dinosaur fossil*

Value

\[ t \text{ (millennia)} \]

“road kill”

“pretty interesting” ->

*what’s more static than a fossil? and inexpensive storage too!
Incremental fossil value?

Value

“road kill”

“I’m more certain” ->

“pretty interesting” ->

$t$ (millennia)
Value of a more complete dinosaur

Value

"road kill"

"pretty interesting" ->

"I’m more certain" ->

"off the charts" ->

Jurassic Park

$n$ fossils

one fossil

$t$ (millennia)
My health information (personal perspective)

cumulative cost

cost / unit time

value to me

Value

time

My health information (personal perspective)

Value

time

cumulative cost

cost / unit time

value to me

$
My health information (aggregate perspective)

Value

- value to me
- value to others
- cumulative cost
- cost / unit time

Time →
What are the constraints on the supply of information?
Constraints on supply

• Consent — *my health is none of your business!*
• Collection — *did I write that down?*
• Veracity — *I really didn’t mean what I said*
• Storage — *where did I put that note?*
• Aggregation — *is all of this really you?*
• Encryption — *has anyone seen my keys?*
• Presentation — *all the King’s horses....*
• Interpretation — *what did you mean by that?*
What are the important steps forward?
A solution framework

1. Incremental efforts
   – time, alignment, money, personnel, will

2. Transformational efforts

3. Intractable problems
   – Scientific / computational
   – Semiotic / hermeneutic
   – Conflicting objectives (privacy / disclosure)

ONC & PCAST: Incremental, transformational, or intractable?
Hypothesis: Think of it as one larger problem

*Three variants* of a single challenge:

1. Acute access for patient care
2. Intermediate aggregation and secondary use
3. Long-term preservation and access

Each variant *must* leverage the others with common approaches to:

- consent, collection, aggregation, interpretation, use
- governance, economic sustainability
Through governance and sustainable models...

- **Consent** — “digital resurrection” through data “living wills”
- **Collection** — leverage EHRs and standards
- **Veracity** — plan for more ephemera, PHR
- **Storage** — keep storage costs low; maintain context
- **Aggregation** — provenance and identity meta-data
- **Encryption** — think “keys” for the long-term
- **Presentation** — incremental models; different abstraction levels
- **Interpretation** — naming authority, versions,
Accept complexity

Figure 1. Complete sequence of the cDNA encoding the bovine growth pre-hormone. The deduced amino acid sequence is indicated and the pre-sequence is underlined. The first amino acid of the mature hormone (Ala) is marked in bold. This sequence is deposited at GenBank (accession number AF034386).
Why this is important

*We become what we behold.*

*We shape our tools,*

*and then our tools shape us.*

*Marshall McLuhan, 1964*
Time value of information: many possibilities

For a given data set from an individual or a population, the time value of information can be:

- A decay curve
- An maturing value curve
- A perishable value curve
- An inverse perishable value curve
- Undecidable or random
Some central questions

• What is the single most important step that has to be taken to enable preservation and sharing or re-use of EHR?

• What is the single greatest obstacle to preservation and sharing or re-use of EHR?