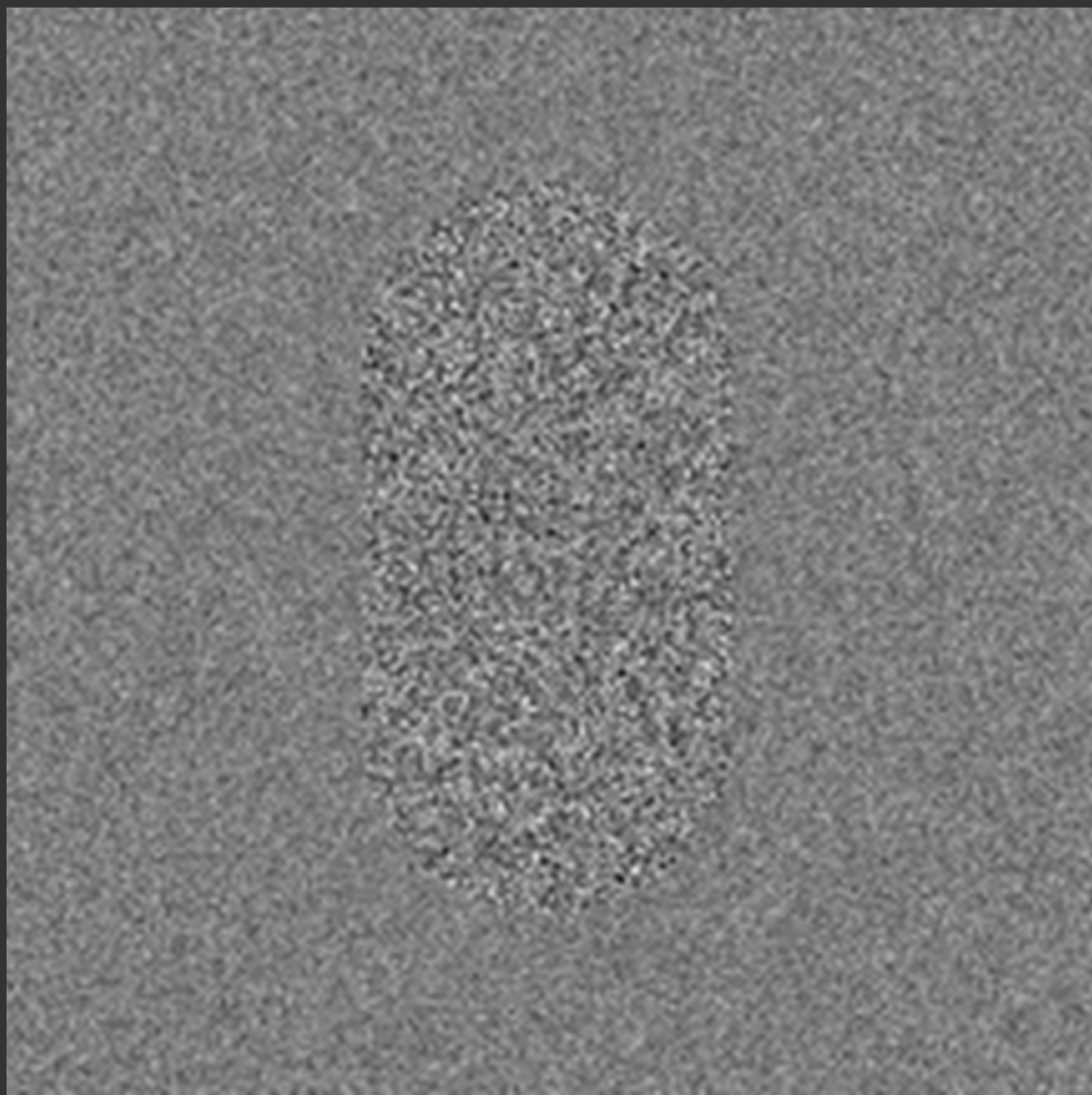


Image Archives: Where are They and What Next? (Why GenBank, but not ImageBank?)

Terry S. Yoo, PhD

OHPCC / LHNCBC / National Library of Medicine
National Institutes of Health





Where are we?

Who am I? Why am I here?

Where are they? Why aren't they here?

- The Infancy of Imaging Informatics
- The rapid rise of new informatics
 - Search
 - Privacy
- Leadership
 - Risk
 - Mission

Some recommended reading...

Koans (paradoxes)...

- It's all just bits.
- Perfection is normal.
- There is want in the midst of plenty.
- Processing is power.
- More the same can be a whole new thing.
- Nothing goes away.
- Bits move faster than thought.

QuickTime™ and a
decompressor
are needed to see this picture.

The Infancy of Imaging Informatics

QuickTime™ and a
decompressor
are needed to see this picture.

- Age of imaging informatics?
- How to parse images? Classify? Keywords?
- Content-based information retrieval?

The rapid rise of new informatics

The rapid rise of new informatics

- Search
- Privacy

QuickTime™ and a
decompressor
are needed to see this picture.

Risks to Privacy

UK Government in Uproar Following Data Loss

Misplaced disks contained personal information on 25 million taxpayers

Nov 19, 2007 | 06:51 AM | [0 Comments](#)

By Tim Wilson
Darkreading

Legislators in the United Kingdom are calling for heads to roll following a series of errors that may have exposed 25 million Britons' personal information.

The entire database of child benefit recipients maintained by [Her Majesty's Revenue and Customs \(HMRC\)](#) department has gone missing after being posted to the National Audit Office by a junior official, according to [reports](#).

The junior official's action was contrary to regulations that govern the handling of data within HMRC, the reports say. The official then compounded the mistake by re-sending the disks when they did not arrive on the first try.

The lost data was password protected, but not encrypted, according to another [report](#). "A criminal could break into these files in a matter of minutes," said Simon Davies, a senior visiting fellow at the London School of Economics who specializes in data security.

There have been at least three major leaks involving HMRC in the past three months, according to British press. In October, a laptop containing data on up to 2,000 investors was stolen. In Parliamentary answers, officials revealed that 41 laptops have been stolen from HMRC in the past 12 months.

A CD carrying details of 15,000 Standard Life customers was also lost by the agency.

Aside from expressing their outrage, British legislators are using the breach as a springboard to renew discussions over the U.K. Data Protection Act, which currently does not require companies to notify either the authorities

Privacy risks are real

- Threatens progress
- Can affect governments
- Bring down an industry giant (Sony)?

A growing experience base...

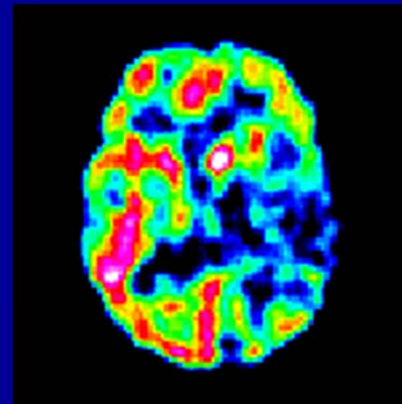
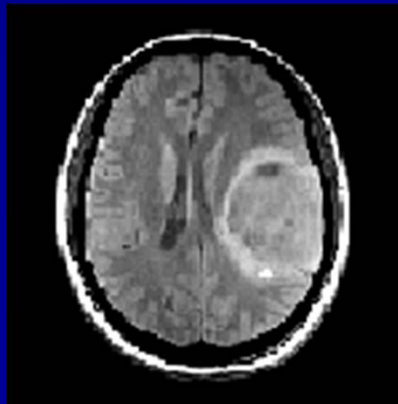
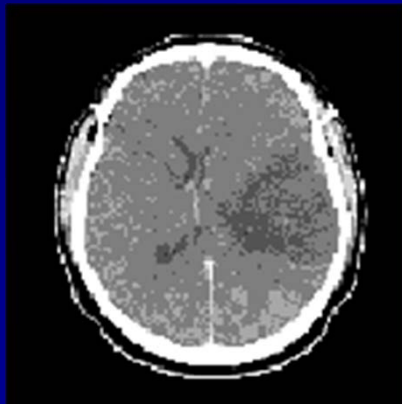
Personal failures

- WRAMC Colonoscopy Collection
- Mayo BIR Image Collection
- Vanderbilt RIRE

Agency experience

- BIRN
- LIDC, RIDER
- caBIG

Retrospective Image Registration Evaluation Project (1995 – present) J. Michael Fitzpatrick, P.I.



- The SAT test for rigid, non-deformable multimodal brain registration.
 - Separate training and test data. Provide tests under careful control.
 - Only scores are returned... no solutions.
- Funded across NIH over several installments.
 - Originally NINDS, then NCI, then NIBIB.

Some institute success...

NLM

OSA

Kitware

- ISP
- ITK
- VolView

QuickTime™ and a
decompressor
are needed to see this picture.

<http://www.opticsinfobase.org/isp.cfm>



Hope...

Image collections for research are coming...

- Phantoms, Give-a-scan, OSA-ISP
- Computer vision
- Leadership
 - Risk
 - Mission

<http://www.cs.cmu.edu/~cil/v-images.html>

QuickTime™ and a
decompressor
are needed to see this picture.

Principles from the First International Strategy Meeting on Human Genome Sequencing

(Bermuda, 25-28 February 1996)

- Primary Genomic Sequence Should be in the Public Domain.
- Primary Genomic Sequence Should be Rapidly Released.
 - Sequence assemblies should be released as soon as possible; in some centres, assemblies of greater than 1 Kb would be released automatically on a daily basis.
 - Finished annotated sequence should be submitted immediately to the public databases.

Imaging for Biological Systems: An Open Science Challenge

- Infrastructure building.
 - Examples: GenBank, ITK
 - Possible Need? Visualization of gene expression.
- Validation, reproducible science.
 - Example: Vanderbilt RIRE
 - Possible Need? Online tools: Databases, Ground Truth, Repositories, validation resources.
- Open Source. Open Data. Transparency.
 - Examples: Genomic Revolution
 - Possible Need? A Bermuda Conference?

Clash of interests?

- Technology Development or Combating Disease?
 - Customer?
NCI, NIAID, NHLBI, NIGMS, NIDDK, NINDS, NIMH, NICHD, NCRR, NIEHS, NIDA, NEI, NIA, NIAMS, NHGRI, NIAAA, NIDCR, NIDCD, NLM, NIBIB, NINR
- Public Good or Self Interest?
 - Example: Vanderbilt RIRE
 - Possible Need? Online tools: Databases, Ground Truth, Repositories, validation resources.
- Open Source. Open Data. Transparency.
 - Examples: Genomic Revolution
 - Possible Need? A Bermuda Conference?

The Downward Spiral:

(taken from Newsletter of the Clinical Molecular Genetics Society 2000)

Book review: The Lord of the Helix by ACG Tolkein

This book, written by an obscure Cambridge academic has all the makings of a classic, tragic fantasy. It tells of three hobbits, WashU, Doe and Sanger meeting the Wizard Wellcome (called NIH in the Lands of Westernesse). Their quest is to discover the ancient helix and prevent the Dark Lord Celauron using it to control the fate of the Middle Earth. After the Council of Elrond (held in Bermuda) the hobbits decide that, with their allies they must take the Helix and cast it into the public domain, but the quest is dogged by the slimy green Ventrums, and his allies the Trolls of Wall Street. Beautifully written and highly evocative of an era gone by when truth was seen as more than just a commodity, I can't wait until the final volume is published.

Some Hope

The Development of the Internet

- Only one key application
- Interoperability was a higher priority than standards
- No one asked about the cost or the eventual market
- Just Do It™

Noam Chomsky

- Single handedly redefined the field of cognitive psychology
- Created ripples throughout science (including computer science)
- Chomsky hierarchy matches the formal study of machine languages (FSM, PDA, Context Free, Turing Complete)
- 100% opaque

A deep challenge

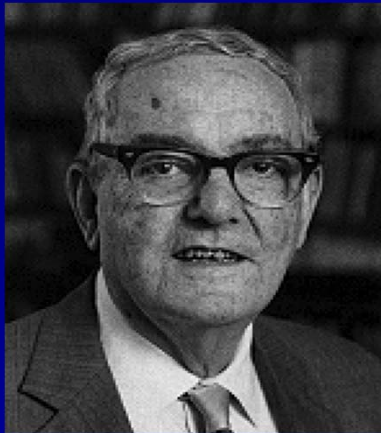
Be a Chomsky

- Study people and the way they organize images
- Learn how to parse images
- Reinvent a field or two

Some Easier Challenges

- Give up your data
 - Overcome Legal constraints
 - Overcome Political constraints
 - Overcome Economic predispositions
- Study the basics
 - Chomsky's example
 - Watch how minds organize images
 - Try to parse images

A Poverty of Attention



“A wealth of information creates a poverty of attention and a need to allocate it efficiently.”

-Herbert Simon,
Mathematician
Noble Laureate, Economics