State of Pathology Informatics: What’s Essential Today and Tomorrow for Success, from Better Collections and Digital Pathology to Synoptic Reporting and Tracking Pathologist Productivity

Ulysses G. J. Balis, M.D., FASCP, FCAP, FAIMBE
Professor of Pathology &
Director, Division of Pathology Informatics
University of Michigan Health System
ulysses@med.umich.edu
www.pathology.med.umich.edu
General Comments

• Concerning the approach to access and use of information:
  – The world is evolving to a *relational* model of informational interchange and yet, our LIS infrastructure remains very much centered on a *transactional* model
• The EMR is not the only access node to the health record
  – Reverse federation is a key solution
• Encoded data in existing lab results can benefit from use of Integrated Diagnostics Architecture
• Clinicians are interested in answers to clinical questions and not to how laboratories compartmentalize *discrete* tests into *discrete* reports
  – Post analytical integration and aggregation of results is needed
  – Exemplar: Rich media molecular reporting
• The consumer is no longer solely the clinician
  – Data must be transformed for both downstream human and machine-based consumers; presently, most of our LIS infrastructure is uninformed of this important and growing need
Part 1: Computational Pathology, Data Analytics and Information Theory

• The emergence of Computational Pathology and Formal Data Analytics for Both AP and CP
• Emergence of the Formally Trained Computational Pathologist
• Encoded Data:
  – Discussion of the limits of human cognition
  – Application of Machine Learning and Analytics to Laboratory Results
  – Computationally-based clinical testing – Multi-analyte Assays with Algorithmic Analysis (MAAAs) with Thiopurine Analyte testing as an exemplar
Data Portfolio:
Contemporary Pathology Setting - 2015

The Present
Data Portfolio:
Digital Pathology Workspace ca. 2020

The Not-too-distant Future
Motivation: Why Develop Image-Based Search Technology In Pathology?

• Find all entries in a repository that match the current image, based on a region of interest of image data itself, and not text or metadata-based search predicates

• Extract associated metadata from returned matching images, providing some measure of equivalence or prediction for the current image (a quantitative class of operation)
  – Diagnosis for returned matching cases
  – Biological potential of malignancy (e.g. survival)
  – Extracted Kaplan-Meyer statistics
  – Historical responses to therapeutic agents / clinical course
  – Association with genomic data already known for the image-matched cohort of cases (in essence, the constitutive image features can become a proxy for previously established multi-dimensional correlates between morphology and the molecular basis of disease
Image-Based Search: A Definition

- Also Known as Content-Based Image Retrieval (CBIR)
- The process of searching libraries of images with exemplar images (or exemplar regions of interest)
- Easily fits in the computationally non-trivial class of search tasks
- Opportunities for both hypothesis-based approaches, as well as generalized (non-specialized) classes of solutions
- Central Hypothesis: Effective deployment of CBIR solutions will be one important gateway to more effective utilization of imagery data (in addition to passive retrieval/view activities)
The availability of digital whole slide data sets represent an enormous opportunity to carry out new forms of numerical and data-driven query, in search modes that are not based on textual, ontological or lexical matching.

- Search image repositories with whole images or image regions of interest
- Carry our search in real-time via use of scalable computational architectures
- Higher order space bioinformatics searches can finally include quantitative histology (e.g. combined search of histology, radiology and genomic repositories offers an significant potential for enhanced statistical power)

Known as Content-Based Image Retrieval (CBIR)

Resultant Heat Map with gallery of matching images and any/all associated diagnostic or decision support data
1. An approach to diagnosis that incorporates multiple sources of raw data (e.g., clinical electronic medical records, laboratory data including "-omics," and imaging [both radiology and pathology imaging]); extracts biologically and clinically relevant information from these data; uses mathematic models at the molecular, individual, and population levels to generate diagnostic inferences and predictions; and presents this clinically actionable knowledge to customers through dynamic and integrated reports and interfaces, enabling physicians, patients, laboratory personnel, and other health care system stakeholders to make the best possible medical decisions.

2. More generally, using computation for the interpretation of multiparameter data to improve health care.
Emergence of the Formally Trained Computational Pathologist / Laboratorian

• With the emergence of increasingly computational and multi-dimensional bioinformatics data that can be constitutive to the rendering of diagnoses, a competent and established cohort of Computational Pathologists will become essential.

• Although molecular pathology trained individuals would appear to be adequate for this role, there are additional skill sets in information theory and data science that are needed
Tools, Approaches and Methods Intrinsic to Computational Pathology

• Tools:
  – “R”
  – MatLab
  – Python, Perl, PHP
  – Docker, Jenkins, etc.

• Approaches
  – Bayesian methods
  – Galois fields
  – Modular Functions
  – Predicate Calculus
  – Quantum Mechanics Notation of primary medical data
  – Graph Embedding
  – Manifold optimization / Dimensional Reduction
  – Machine Learning
    • Random Forest (in R)
    • Boosting

• Methods
  – Validation Strategies
  – Data Orchestration Models
  – Computational pipelines
  – LIS integration Strategies
  – Delivery vehicles (Web portals and Web Services model)
Emergence of Data Analytics as an Integral Function of the LIS

• Observation: Increasingly, the lab is being asked to provide three classes of results that require a level of sophistication that has not yet been uniformly realized:
  – Integrated reports that span multiple discrete orders (e.g. flow cytometry, hematology, anatomic pathology molecular analysis as a single integrated leukemia/lymphoma report
  – Computationally extracted results from primary data, requiring curated analytical algorithms and statistical methods
  – Dynamically regenerated reports, pushed to the original ordering provider, upon the availability of new actionable medical knowledge (the prototypic example being an uncategorized variant changing to a “must-call” variant, based on newly reported scientific data)

For each of these three new modes of reporting to take place, new programing logic and external interfaces from the LIS will need to be developed and maintained.
Mining for Encoded Data / Recognition of the limits of human cognition

• The Human cognitive process excels in interpretation of discrete data
• Knowing how to interpret the data is also critical (the pipe example), and this information is not always available, *a priori*
• As complexity level of data increases, two things happen:
  – Total possible classes of orthogonally-encoded data grows geometrically with respect to the linear size of the baseline data.
  – Human potential to extract such encoded data diminishes with the degree of dimensional growth

Net result of increasing data complexity without computational assistance: Incomplete Extraction of Medically Actionable Information
Hypotheses

• Distinct and medically actionable **diagnostic Information** may be distributed throughout a multi-dimensional parameter space, in addition to the obvious discrete information represented by each atomic data element

• The consistent extraction and interpretation of such data may be **unreachable** by the limits of human cognition, owing to both complexity and subtlety of its distribution
Types of *Data Adjacency* that can easily exceed human cognition:

- Parallel
  - Aggregated
  - Spatially encoded
- Time Series
- Variegated data sets
- Distributed
  
  Any combinations of the above
Increasing Complexity

Single Analyte

Chem 7

1 data element: Simple linear inference model

7 data elements: limit of experiential threshold of encoded data extraction

28 Data Elements: encoded data present beyond human cognitive limit

250 Data Elements: encoded data including prognostic data likely present

10^3 - 10^4 data elements

10^5 data elements

10^8 data elements

10^11 data elements

10^12 data elements

10^20 data elements

NGS Time-Series Data

Population-Level NGS Time-Series Data

NGS Time-Series Data Set

Library of Whole Slide Images

Single Whole Slide Image Data

Expression Array Data

Tissue Microarray Study

Expression Data

Time-Series Routine Lab Studies

Comprehensive Chemistry + CBC

Supercomputing Threshold

Threshold for complete cognitive data extraction
Exploring The Entropy / Image Content Curve

<table>
<thead>
<tr>
<th>Type of Information</th>
<th>1986</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>2.6</td>
<td>295</td>
</tr>
<tr>
<td>Broadcast</td>
<td>432</td>
<td>1900</td>
</tr>
<tr>
<td>Telecommunications</td>
<td>0.281</td>
<td>65</td>
</tr>
</tbody>
</table>

Logarithmic Magnitude of File Size / Data Content
In the setting of the larger context of personalized and precision diagnostics, the data adjacency problem becomes daunting to detection by human cognition alone, if encoded data is indeed present.

From: Athey and Omenn, 2010
Assessing presence of absence of disease states distributed between a multitude of deviated analytes is not trivial in the setting of the data being encoded among many discrete data elements.
Exemplar:

**Multi-Analyte Assays with Algorithmic Analysis (MAAAs)**

- Our existing routine laboratory results data may contain many levels of “encoded” additional information, with this data being: real, distinct, diagnostic, prognostic and theranostic.
- At present, there is no systematic approach to:
  - structure the capture of such data in the LIS itself
  - Implement computational solutions and rule sets
  - easily reproduce such capture across disparate LIS architectures
  - Obtain constitutive input data electronically from requesting laboratories
  - forward integrative results to downstream repositories (e.g. data consumers) in a manner that preserves semantic layers of information
ORIGINAL ARTICLES—ALIMENTARY TRACT

Algorithms Outperform Metabolite Tests in Predicting Response of Patients With Inflammatory Bowel Disease to Thiopurines

AKBAR K. WALJEE,† JIELI C. JOYCE,‡ SUJAN WANG,# ADITI SAVINEN,* MARGARET HART,* JI ZHU,† AND PETER D. R. HOGINS*†

†Division of Gastroenterology, Department of Internal Medicine, ‡General Clinical Research Center, and Department of Statistics, University of Michigan, Ann Arbor, Michigan; and #Department of Statistics, University of Wisconsin, Madison, Wisconsin

BACKGROUND & AIMS: Levels of the thiopurine metabolites 6-thioguanine nucleotide (6-TGN) and 6-methylmercaptopurine commonly are monitored during thiopurine therapy for inflammatory bowel disease despite this test’s high cost and poor prediction of clinical response (sensitivity, 62%; specificity, 72%). We investigated whether patterns in common laboratory parameters might be used to identify appropriate immunologic responses to thiopurine and whether they are more accurate than measurements of thiopurine metabolites in identifying patients who respond to therapy. METHODS: We identified 774 patients with inflammatory bowel disease on thiopurine therapy using metabolite and standard laboratory tests over a 24-hour time period. Machine learning algorithms were developed using laboratory values and age in a random training set of 70% of the cases; these algorithms were tested in the remaining 30% of the cases. RESULTS: A random forest algorithm was developed based on laboratory and age data; it differentiated clinical responders from nonresponders in the test set with an area under the receiver operating characteristic (AUROC) curve of 0.656. In contrast, 6-TGN levels differentiated clinical responders from nonresponders with an AUROC of 0.584 (P < .001). Algorithms developed to identify thiopurine nonresponse (AUROC, 0.813) and thiopurine shunters (AUROC, 0.757) were accurate. CONCLUSIONS: Algorithms that use age and laboratory values can differentiate clinical response, nonadherence, and shunting of thiopurine metabolism among patients who take thiopurines. This approach was less costly and more accurate than 6-TGN metabolite measurements in predicting clinical response. If validated, this approach would provide a low-cost, rapid alternative to metabolite measurements for monitoring thiopurine use.
An Exemplar of Encoded Data Use in Diagnostics:
Computationally-Based
6-MP Testing

- Entirely computational, based on previously ordered tests (CBC & Comprehensive chemistry panels)
  - No actual bench testing
- *Better* predictive power than the molecular bench-based test
- Heavy computational burden for initial validation and continued performance certification
- Adds ability to predict
  - patient compliance
  - detection of metabolic shunting (toxicity)

*An ideal candidate for cloud-based LIS computing*
Figure 2. Algorithm accuracy and predictor importance of random forest machine learning models based on common laboratory tests. (A) AUROC for clinical response with the random forest clinical response algorithm and with 6-TGN alone. (B) Predictor importance for the random forest clinical response algorithm. (C) AUROC for nonadherence with the random forest nonadherence algorithm. (D) Predictor importance for the random forest nonadherence algorithm. (E) AUROC for shunting to 6-MMP with the random forest shunting algorithm. (F) Predictor importance for the random forest shunting algorithm.
A rules-based clinical algorithm: the ideal situation for automation in clinical reporting
Variable Importance of Each Analyte for an Overall Clinical Response Algorithm
Comparing Machine Learning Algorithms to 6-TGN

ROC curves for predicting BR

- 6TGN: AuC = 0.831, 95% CI = [0.820, 0.842]
- V1: AuC = 0.576, 95% CI = [0.561, 0.692]
- V2: AuC = 0.470, 95% CI = [0.384, 0.556]
The ThioMon Test in Actual Clinical Use

- Clinical Order for CBC and Comprehensive Chemistry (written or electronic)
- Bench Testing (<3 hours)
- Random Forest Algorithm on R Server: 2.1 seconds
- Decision Support within minutes
Present State of Reference Lab Thiomon Testing

- With the current service model, external patient blood specimens must be physically transported to the U-M clinical Lab
- This is inefficient
- A better model would be to simply transport the CBC and Chemistry *electronic results data* (concept of eTubes)
- This required a split study to demonstrate equivalence
A Split Study to Demonstrate Utility of ThioMon Testing via the eTubes Concept

Clinical Order

Remote Pathology Clinical Laboratory

Initial Lab Results from External site electronically transferred to the central eTubes server

UMHS Pathology Clinical Laboratory

Initial Lab Results from local lab electronically transferred to the central eTubes server

Final Comparison of Computed Results Between Remote and local Locations

2.1 seconds
ThioMon Testing: Realized Features and Benefits

• More predictive than the current “Gold Standard” test (Prometheus Labs), effectively making this the “new Gold Standard test”

• Immediate resulting capability with simply CBC and comprehensive chemistry panels as input data (no additional bench testing needed)

• Web Infrastructure combined with the eTubes concept makes this technology available on a global basis, without transport of specimens

• Operational resources required are very modest
  – R server (virtual machine)
  – Web Portal (virtual machine)
  – 0.1 FTE for website and data pipeline maintenance
How ThioMon Works

• Integration with LIS orders and primary results
  – CBC and Comprehensive Chemistry results constantly monitored for new paired entries
• Paired Results reflexively order a ThioMon Computational Assay
• The Primary results are looped to an LIS-attached R server
• 11,500 optimized / validated random forest rules on the R Server generate a Computational Report
• The Computational report is returned to both the LIS and to a separate Web Server / Web Services Layer
• The clinician is then able to review the result
Current Use

- In use for past 3 years
- Implemented in the SCC-Soft Lab Information System
- Fully interfaced to the local EMR (Epic MiChart)

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>SHUNT</th>
<th>CLINRESP</th>
<th>NONCOMP</th>
<th>THIO INTRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/05/2010</td>
<td>1.4</td>
<td>80.6</td>
<td>0.4</td>
<td>GOOD: see text</td>
</tr>
</tbody>
</table>

**THIO INTRP**
Interpretation: Assuming this patient has been on the same dose of thiopurine medication for at least 4 weeks and is at steady state: This patient has had a good hematologic and chemistry response to thiopurines, has a low probability of shunting, and a low probability of noncompliance. If this patient has not obtained a good clinical response, they may have a non-inflammatory cause of symptoms, or if inflamed, may need a different form of therapy, as dose increases are unlikely to produce large therapeutic gains.
Thiomon results, now rendered as a webpage:
Initial Findings With 3.5 Years of Use:

- Performance of ThioMon greatly exceeds the bench-based molecular test, in terms of predicting biologic response
- Adoption of use has increased by 100-fold since the third publication
- Now possible to deploy the test as an orderable item by other institutions, as a web-based, eTubes test
Application of Machine Learning and Analytics to Laboratory Results

• Recognition that there are at least several thousand additional encoded results in simply the CBC/Comp pairing
  – Hepatocellular CA early prediction
  – Hepatitis C to hepatocellular CA progression prediction as a stratification tool for tx for Hepatitis C
  – C Diff colitis prediction upon ED presentation
  – Many more.....

• Expand use of Computational Pathology methods to all fields of primary data (eg histology, etc.)
Part 2: LIS Architecture and Implementation

– Observations
– Widespread adoption of Cloud-based Applications
– Emergence of web-based (HTML 5.0) LIS architectures
– Emergence of Rapid Prototyping Tools
– Interoperability of Clinical Lab Results across disparate systems and locations
Part 2: LIS Architecture and Implementation

• Observations:
  – Now nearly 45 years old, LIS solutions have greatly matured in some respects, and yet remain entirely under-developed in key areas:
    • Interoperability between disparate systems still a dream
    • Urgent need for a consistent and unified data representational model
    • Usability and User Experience (UX) engineering of LIS systems is far from what has been achieved in other IT sectors
    • Reliability and uptime lag behind other enterprise systems, such as the EMR, owing to antiquated architecture and software development/validation methods
Cloud-Based Applications

• Motivation:
  – Prior to cloud based solutions, application availability and reliability were largely a function of the robustness of local IT infrastructure. Economies of scale make it cost-prohibitive for many enterprises to implement a Tier 4 data center.
  – By leveraging cloud-based solutions for both data storage and servers, it is possible to benefit from the redundancy afforded by multiple Tier 4 data centers.
Comparison of Deployment Models

Conventional Local LIS

- **Advantages:**
  - Fewer external dependencies such as network connectivity
  - Somewhat reduced risk of unauthorized PHI disclosure
  - Well-defined enterprise IT policies concerning governance and data security

- **Disadvantages:**
  - Much higher TCO
  - Greater risk of service interruptions / data loss
  - Impeded ability to scale with growth
  - Need for local IT expertise along multiple skill sets

Cloud-based / Remote-Hosted

- **Advantages:**
  - Much lower overall TCO
  - Superb availability / data permanence
  - Geospatial vigor

- **Disadvantages:**
  - Absolute dependence on wide area network availability
  - Potential for data breach if both remote and local systems are not properly secured
  - *Enterprise data security policies may not yet have recognized cloud solutions as an acceptable model*
Cloud-based Applications

• Enabled Operational Model
  – Selection of a vendor that provides for true HIPAA compliance and transactional auditing at the network abstraction layer
  – Multiple, distributed sites is a requirement
  – Consideration of the addition of minimum viable local IT functionality, in the unlikely but conceivable setting of Internet connectivity failure
    • Need for both storage and server infrastructure, locally
Cloud-Based Applications

*(General Architecture)*

- No longer should be viewed as an “extreme” or “aggressive” IT strategy
- Multiple vendors for HIPAA-compliant data storage and server infrastructure
- Increasing number of LIS and Middleware vendors now offer cloud-based solutions, in the form of Software as a Service (SaaS).
- A useful byproduct of this approach is the simplified ability to use multiple form-factor devices as LIS access points (responsiveness).
Emergence of Web-based (HTML 5.0) LIS Solutions

• Motivation: The plurality of web browsing hardware, form factors and operating systems creates a challenge for maintaining the stability of the LIS thick client software, in the setting of ever-present updates and patches to the enterprise operating system.

• Case reports of entire laboratories being temporarily blocked from their LIS due to the software no longer being able to run properly.

• Need for responsiveness to support multiple form-factors
Emergence of Web-based (HTML 5.0) LIS Solutions

• Opportunity: HTML 5.0 based applications run entirely within the Web browser framework, and are thus much less susceptible to failure due to modification of the host environment. Additionally, they can made to be hardware-independent, adjusting the user interface to match the form-factor of the device (with the term of art being “Responsive”)
Emergence of Rapid Prototyping Tools

• Observations:
  – LIS vendors rarely generate an optimal user interface with initial software versions. There are many reasons for this:
    • Incorrect assumptions
    • Insufficient (or no) vetting of design with users
    • Underpowered development team in terms of UX engineering
  – New tools are now available that allow the end user to rapidly prototype a functioning wireframe user interface, greatly accelerating application development
# The Pillars of a Container Centric Infrastructure

<table>
<thead>
<tr>
<th>Source Control</th>
<th>Build System</th>
<th>Image Repository</th>
<th>Container Cluster</th>
<th>Service Discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>GitHub</td>
<td>Jenkins</td>
<td>Docker Registry</td>
<td>Hosts + Deploy Scripts</td>
<td>etcd</td>
</tr>
<tr>
<td>GitLab</td>
<td>GitLabCI</td>
<td>Docker Trusted Registry</td>
<td>Swarm</td>
<td>Zookeeper</td>
</tr>
<tr>
<td>BitBucket</td>
<td>Bamboo</td>
<td>Docker Hub</td>
<td>Mesos</td>
<td>Consul</td>
</tr>
<tr>
<td>Stash</td>
<td>TravisCI</td>
<td>Quay.IO</td>
<td>Kubernetes</td>
<td>Vanilla HAProxy</td>
</tr>
<tr>
<td>Gerrit</td>
<td>CircleCI</td>
<td></td>
<td>Rancher</td>
<td>DNS</td>
</tr>
</tbody>
</table>

*Note: Many service discovery options must be coupled with something that can route (HAProxy).*
Gitlab

- Open source
- Close-enough mirror to the familiar (GitHub)
- Low barrier to entry for non-developers
- Existing integration into other “pillars” (Jenkins)
- Now comes in a handy dandy container
  - “docker pull gitlab/gitlab-ce”
  - https://hub.docker.com/r/gitlab/gitlab-ce/
Jenkins

- Open Source
- Time tested solution
- Plug-ins galore
  - Docker
  - Mesos Cluster
  - Gitlab Webhooks
- No real constraints on architecture
  - No plugin? Use Bash!
- Blank slate for any project
- Windows Support
Docker Trusted Registry

- **Security Requirements**
  - 100% on premise.
  - LDAP backed.
  - RBAC - v0 api available, use and integration capability growing.
    - [https://docs.docker.com/docker-trusted-registry/api/](https://docs.docker.com/docker-trusted-registry/api/)
  - One-click upgrade.
  - User friendly UI.

- **Support Offerings**
  - 24/7 support REQUIRED for use in a clinical care setting.
  - Guided install with staff on-site where needed
Mesos (Cont.)

Just to name a few....

- https://open.mesosphere.com/frameworks/
Interoperability of Clinical Lab Results Across Disparate Systems and Locations

• Motivation: There is significant unmet need to transport electronic lab results to external interests, and yet, there is no single unified standard

• In the present reality, every request for such electronic data is addressed by the development of a custom interface (typically HL7)

• The cumulative opportunity cost of developing and then maintaining these interfaces is substantial
Interoperability of Clinical Lab Results Across Disparate Systems and Locations

• Opportunity: Utilize a single standard to transport all classes of lab data

• Observation: Two factors are accelerating the creation of a solution:
  – Health Information Exchanges
  – Requirements, as stipulated by Meaningful Use, that health organizations must be able to forward electronic lab results to outside interests and information exchanges
Interoperability of Clinical Lab Results Across Disparate Systems and Locations

• Likely Implementation Models
  – CDISC: Clinical Data Interchange Standards Consortium
    • An emerging data model for electronic interchange of lab data, including properly encoded reference intervals and units.
    • Includes a reference schema, making the generation of interoperable results messages a far simpler task
Part 3: Adoption of Precision Medicine

- Federation of electronic health data at the enterprise level and reverse federation in support of Precision Laboratory Medicine
- Personalized Reference Intervals
- Time Series Reporting
- The Integrated Diagnostics Delivery Model
Federation of Electronic Health Data at the Enterprise Level and Reverse Federation

• Recognize that central IT departments have increasing demands and often diminishing composite resources to meet such demands

• Recognize that ancillary departments are usually the most qualified domain experts in terms of appropriate stewardship of data.

• Recognize that it takes less effort and imparts less risk to store every data element once and no more than just once (affirming the use of SSOT principles)

• Reduce time-consuming rework associated with cascaded interface changes, in the setting of evolution in both workflow and data model practices.
Multiple unique instances of customized HL7 interfaces which are not necessarily centrally managed or organized for consistency in deployment style.
A Conventional EHR Data Model

- User
- Centralized EHR
- LIS
- RIS
- OR
- E.D.
- other

- Conventional HL7 interface
- Browser or Thick Client EHR Viewer
...vs. a Federated Data Exchange Model

Consequences of shifting to a SQL-based SSOT model:
- Data only represented once in overall enterprise model
- Reduction in number of interfaces requiring support
- Potential to transfer classes information other than text
- Reduction in support responsibilities of central hospital IT.

SSOT is shifted to the appropriate domain-specific stewards of data, as opposed to being in the EHR domain.
Concept: Reverse Federation

- Not only does the LIS serve the greater EHR connectivity ecosystem, so too should the greater plurality of EHR repositories serve the LIS and its emerging workflow needs, in the capacity of decision support data feeds assisting in report generation.
  - Increasingly essential for:
    - Molecular reporting
    - Personalized medicine
    - Synoptic cancer checklist reporting of correlation of histopathology staging in concert with clinical stage and longitudinal reporting (AJCC Cancer Staging Manual, 7th edition and later...)
What is Reverse Federation?

Potential Revised Data Model

Consequences of shifting to LIS-centric world view model:

- Data transfer to LIS is ephemeral – only kept as long as needed for display
- LIS user is not required to expend additional effort to log on to multiple clinical systems to access all required clinical data
- Likelihood that clinical data is accessed whenever it should be is greatly increased, thus driving improvements in consistency and quality of laboratory reporting
Reverse Federation – SWOT Analysis

• Strengths
  – Pathologists and laboratorians can more effectively review the complete medical record from within the LIS, saving time and improving quality

• Weaknesses
  – Completing a reverse federated feed from the EMR remains a technically challenging process

• Opportunities
  – Create reimbursement models for hosting such value-added workflow

• Threats
  – The “Primary EHR solution” trend now under way at many health enterprises may remove this architectural approach as a possibility, owing to the monoculture of the industry
High Level Workflow Map

1: Inbound Lab Orders
2: Clinical Case History / Images
3: All Clinical Lab Data
4: Barcode Tracking Data
5: Return of Complete Surgical Pathology Reports from the cockpit

EPIC Chronicles
EPIC Resolute (Billing)
EPIC ADT/Prelude Registration
Other Clinical Repositories
DICOM-Based PACS

Interface Engine (WBI, Cloverleaf, eGate, etc.)

Digital Workflow Engine
Application Server

Overall Application Suite

LIS
Barcode Tracking System

High-Performance Scanner Server with Local Image Store
Mirrored High-performance On-line and Near-line Image Storage

High-Throughput Scanning Facility

Pathology Signout Cockpits

1
2
3
4
5

1: Inbound Lab Orders
2: Clinical Case History / Images
3: All Clinical Lab Data
4: Barcode Tracking Data
5: Return of Complete Surgical Pathology Reports from the cockpit

DICOM-Based PACS

Overall Application Suite

Pathology Signout Cockpits
Personalized Reference Intervals

• A recent series of publications confirms the longstanding hypothesis that for some analytes, the reference interval should be tied to the individual and simply to the local reference population

• Carrying out this type of study requires substantial data from thousands of normal subjects, along with advanced statistical analysis

• New computational approaches make this possible:
  – Random Forest method
  – Reiterative Imputation
Personalized Reference Intervals For Platelet Count Reduce The Number Of Subjects With Unexplained Thrombocytopenia

Carlo Zaninetti, Ginevra Biino, Patrizia Noris, Federica Melazzini, Elisa Civaschi, Carlo L. Balduini

Haematologica January 2015 : Doi:10.3324/haematol.2015.127597
Derivation of gender and age-specific reference intervals from fully normal Japanese individuals and the implications for health screening

Minoru Yamakado a, Kiyoshi Ichihara a, Yoshiyuki Matsumoto c, Yoshiki Ishikawa d, Kiminori Kato e, Yusuke Komatsubara c, Norihide Takaya f, Shohken Tomita g, Reo Kawano b, h, Keisuke Takada b, Kiyoaki Watanabe i

a Department of Nursing, Ashikaga Institute of Technology, 268-1 Ohmoe-cho, Ashikaga, Tochigi 326-8558, Japan
b Department of Laboratory Sciences, Faculty of Health Sciences, Yamaguchi University Graduate School of Medicine, 1-1-1 Minamikogushi, Ube 755-8505, Japan
c National Federation of Health Insurance Societies, 1-2-4 Minamiyoyama, Minato-ku, Tokyo 107-8558, Japan
d Mizuho Health Insurance Society, Otsuuchi Medical Center, 1-5-5 Otsuuchi Chiyoda-ku, Tokyo 160-8176, Japan
e Department of Laboratory Medicine and Clinical Epidemiology for Prevention of Noncommunicable Diseases, Graduate School of Medical and Dental Sciences, Niigata University, 1-757 Asahimachi-dori, Chuou-ku, Niigata 951-8510, Japan
f Medical Corporation Doyukai, 1-15-10 Nishihata, Burumyo-ku, Tokyo 113-0024, Japan
g Kansai Occupational Health Association, 3-1-1 Umeda, Kita-ku, Osaka 530-0001, Japan
h Clinical Research Center, Yamaguchi University Hospital, 1-1-1 Minamikogushi, Ube 755-8505, Japan
i Tokyo Center for Laboratory Medicine, 6-3-17 Shimomeguro, Meguro-ku, Tokyo 153-0064, Japan
From the work of Ichihara, it can be seen that there are three levels of reference Intervals that need to be considered and represented in the LIS:

- Adjusted to the overall local population
- Adjusted to age and gender
- Individually determined and reported, based on imputed normal values
Time Series Reporting with BCR/ABL1 as the Exemplar

- Pain Points
  - Conventional LIS Architectures are not designed or optimized for time-series reporting
  - Need for textual reports, rich media (log-linear plots), and machine consumable results output modalities (concurrently)
  - Need to integrate single patient identity across multiple reference client sources to realize true lifelong reporting continuity
  - Need for orchestration within the lab (multiple instruments) and with external repositories
    - EMR
    - Health Information Exchanges
    - Patient Portals
    - Research Data Repositories
  - Consent and de-identification / anonymization data representation needs to be tightly threaded to all data
  - Solution: use of an Integrated Diagnostics Server infrastructure, in addition to the core LIS
Time Series Reporting with BCR/ABL1

Collection Date

- 8/16/2012: 959
- 8/17/2012: <85.4
- 8/18/2012: <58.8
- 8/20/2012: Not Quantifiable
- 8/21/2012: Inadequate Specimen
- 8/23/2012: 0.643
- 8/24/2012: Not Detected
- 8/25/2012: <0.00933
Integrated Diagnostics Server Delivery Model:
A High Level Architecture

- LIS remains as the anchoring Schema
- Wrap-around functionality enabled by service-oriented architectures and looped messaging
- Parallel human-readable and machine-readable pathways
- Use of Federated architectures whenever possible
- Natural Language Processing capability to be tightly integrated into the overall solution, given the universality of narrative text in need of classification
High Level Integrated Diagnostics Architectural Map

Enterprise Service-Oriented Architecture Message Bus

- Multiple Clinical Data Sources
  - Epic
  - Sunquest (LAB)
  - Sunquest-AP (CoPath)
  - PACS
  - Cancer Registry

Discreet Numerical Data Parsing Pipeline
- Staging
- Numerical Validation
- Relational DB
- Final Data Transformation

Free Text NLP Parsing Pipeline
- Staging
- Lexical Validation
- Relational DB
- Final Data Transformation

Image Scanning Pipeline
- Scanning Center
- Image Aggregation
- Relational DB
- Image Analysis / Informed Detection

Multi-Axial Edge-Connected And Relational Database with High-Performance Cluster

Multiple User Classes
Integrated Diagnostics Server Delivery Model:  
*Resultant Benefits*

- Federation of salient data, as one view by a single expert (or a panel of experts) allows for the synthesis of a diagnostic report that spans discrete test and procedure orders, allowing for:
  - Internal consistency of the final rendered diagnostic data
  - Simplified review of all underlying diagnostic evidence (especially helpful at tumor boards and as time passes)
  - Enhanced assurance that all constitutive diagnostic elements are considered in the formulation of the final diagnosis
  - Enhanced collaboration across disciplines (e.g. Radiology/Pathology) in rendering a final diagnosis
Part 4: Lab Workflow, Automation and Management

- Emergence of widespread adoption of 2D barcoding for lab specimen tracking and workflow
- Universal Specimen Identifiers / Label Interoperability
- Emergence of RFID
Emergence of widespread adoption of 2D barcoding for lab specimen tracking and workflow

• A 2D standard for laboratory specimen labels is under development by CLSI

• Key features will include:
  – Robust error correction in the 2D format
  – Expanded user fields to included additional site-specific data
  – Unique specimen identification to the patient/site level, avoiding the need to relabel samples, upon their forwarding to outside reference labs
Schematized Color Rendering: Optional Vertical Priority Zone at Left

- Patient Identifier Area
- Patient Last, First, Middle Initial
- Unique Identifier
- Date of Birth, Age, Sex
- Bar Code Area
- Bar Code Human-readable Area (accession number*actual barcode content, if different)
- Collection Site, Date/Time, Collector ID
- Optional Site-specific Data Area 1
- Optional Site-specific Data Area 2
- Fixed-width Bar Code Quiet Zone
- Variable-width Bar Code Quiet Zone
- Priority Status and Container Type
Schematized Color Rendering:
Without Optional Vertical Priority Zone

<table>
<thead>
<tr>
<th>Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Identifier Area</td>
<td>Collection Site, Date/Time, Collector ID</td>
</tr>
<tr>
<td>Patient Last, First, Middle Initial</td>
<td>Optional Site-specific Data Area 1</td>
</tr>
<tr>
<td>Unique Identifier</td>
<td>Optional Site-specific Data Area 2</td>
</tr>
<tr>
<td>Date of Birth, Age, Sex</td>
<td>Fixed-width Bar Code Quiet Zone</td>
</tr>
<tr>
<td>Bar Code Area</td>
<td>Variable-width Bar Code Quiet Zone</td>
</tr>
<tr>
<td>Bar Code Human-readable Area</td>
<td>(accesion number*actual barcode content, if different)</td>
</tr>
</tbody>
</table>
Dimensions of Label Elements: With Optional Vertical Priority Zone

Label Dimensions: With Vertical Priority/Container Option

All dimensions in inches / (mm)
Dimensions of Label Elements: Without Optional Vertical Priority Zone

Label Dimensions: Without Vertical Priority/Container Option

- **Total horizontal label extent**: 2.0 (50.8) inches
- **Total printed extent**: 0.917 (23.3) inches
- **Total vertical label extent**: 0.394 (10.0) inches
- **Barcode text height**: 0.12 (2.54) inches
- **Quiet zone**: 0.25 (6.35) inches
- **Barcode symbology zone**: 1.395 (35.4) inches

**All dimensions in inches/ mm**
Example Rendering of AUTO12-A Label With Optional Priority Zone

Ranganathan, Shiyali R
6455765433431 09-AUG-1892 116y M

Col: 05-MAY-2009 17:43 Site:L Arm  byID0047
Comp. Chem, Mg, Osm, Li
Example Rendering of AUTO12-A Label Without Optional Priority Zone

Ranganathan, Shiyali R
6455765433431 09-AUG-1892 116y M

HM-09-132-00234*0000043456753

Col: 05-MAY-2009 17:43 Site: L Arm bylID0047
Comp. Chem, Mg, Osm, Li
Universal Specimen Identifiers / Label Interoperability

• Observation: At present, it is almost universally recognized that a specimen sent to a reference lab for testing will need to be relabeled upon receipt, to allow for compatibility with local automation and tracking solutions in place.

• Opportunity: Utilize both formatting standards and a central secretariat to allow for the original label to be used by the reference lab.
Universal Specimen Identifiers / Label Interoperability

• Solution: Leverage one of the intrinsic aspects of the anticipated AUTO-14 label: The unique site identification field in the overall field sequence.

• No two labels from can ever be the same, with this observation holding true across all sites that will comply with the Auto-14 standard.

• At present, there are initial discussions under way as to how unique site location codes will be assigned, but a likely outcome will be the use of the CLIA license number.
Emergence of RFID Technology

• What is RFID?
  – Radio Frequency Identification
  – A micro chip programmed with a unique serial number can be paired with a simple transmitter to provide positive identification, upon interrogation
Recent Barcode / RFID developments

- Modern tags exhibit much lower failure rate
- Ability to read multiple assets at once (batch mode)
- Active RFID is maturing
- Detection devices are both more reliable and can operate over greater distances
  - e.g. Door portals
  - Fixed-mount and hand-held devices that have dual RFID / Barcode capability
Summary

• Information Technology is accelerating in its deployment in AP and CP

• The greatest challenge remains the need to train and maintain an adequate cohort of specialists to effectively deploy and then maintain this technology

• If properly utilized, the collection of technologies described today will greatly expand our field’s capacity to increase testing volume and accuracy, without staffing level increases.