IT solutions for imaging biomarkers in biopharmaceutical research and development

M. Hehenberger
A. Chatterjee
U. Reddy
J. Hernandez
J. Sprengel

Biomarkers are indicators of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention. The biopharmaceutical industry is building significant molecular-imaging capabilities and in this context is incorporating biomarker concepts throughout its research. In this paper, we discuss and propose information technology (IT) standards and architectures that support incorporation of imaging biomarkers into the drug discovery and clinical development process. In particular, we cover various uses of emerging imaging technologies in biopharmaceutical research and development, examples of imaging biomarkers in therapeutic areas, IT requirements related to the use of imaging technologies, challenges related to the integration of imaging biomarker data with clinical and genotypic data, and the need to integrate external public data sources. We discuss IT standards and architectures associated with the inclusion of biomarker-related data in the submission of new drug applications, with emphasis on imaging technologies. We suggest extensions to the Study Data Tabulation Model of the Clinical Data Interchange Standards Consortium and the JANUS Data Model of the Food and Drug Administration with data elements based on imaging biomarkers.

INTRODUCTION

The biopharmaceutical industry is currently confronted with many challenges, including evolving business models and a lack of productivity in research and development (R & D).[1,2] The conventional "blockbuster" business model wherein "one size fits all" drugs generate enormous profits will eventually have to give way to a new model of targeted treatments.[3]

The current discovery model for pharmaceutical R & D is based on a clear separation of phases, such as target identification and validation (the biological phase), lead identification and validation (the

chemical phase), and preclinical and clinical development. In this paper, we present aspects of the IT infrastructure for a newly emerging R & D model, based on a biomolecular understanding of disease mechanisms and pathways and the use of biomarkers throughout the R & D process.

A biological marker ("biomarker") is defined as a "characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention."[4] Biomarkers related to measurements that provide information about the efficacy and safety of drug candidates are believed to hold the promise of increased productivity for biopharmaceutical research and development. In its Critical Path Initiative,[5] the Food and Drug Administration (FDA) has attempted to guide the industry toward the use of biomarkers that will address efficacy and safety issues and increase research and development productivity. In addition, the FDA has recently introduced new standards regarding new drug submission data, including guidance documents related to genomic and imaging data.

It is expected that biomarker-based drug development will enable better and earlier decision making and that genomic biomarkers will pave the way toward targeted therapeutics. *Surrogate endpoints* are biomarkers that are intended to substitute for clinical endpoints (i.e., characteristics or variables that reflect patients' feelings, function, or survival). They are expected to predict clinical benefit or harm based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.[6]

Imaging biomarkers have received particular attention because of the noninvasive nature of imaging technologies and the obvious link to diagnostic procedures and clinical care. Imaging technologies are increasingly used as core technologies in biopharmaceutical research and development, both in the preclinical and clinical phases of the research and development process. The first introduction of imaging technologies into pharmaceutical research and development happened in the 1980s, as a technology to support animal studies.[7,8] In the preclinical phase, drugs are tested in animal experiments to establish their efficacy and toxicity before moving to clinical trials in patients.

Today, the use of imaging is growing significantly and is generating a volume of data that is taxing existing IT (information technology) infrastructures. Noninvasive imaging has evolved from visualization of tissue anatomy using structural imaging approaches (X-ray and MRI [magnetic resonance imaging]) to a technology platform that comprises multiple imaging modalities and provides information on tissue morphology, tissue physiology, and metabolic as well as cellular and molecular processes. Molecular imaging can be used to study gene expression or the function of gene products (pathway imaging) in a quantitative manner in the intact living organism. This involves advanced imaging techniques (MRI, optical tomography, tissue modeling) as well as the development of specific biological assays for monitoring the presence of a specific target or of a molecular interaction (e.g., a protein-protein interaction). The ability to study molecular events noninvasively, within their full biological context, is contributing to the understanding of the normal and diseased organism.[9,10]

Since the 1990s, imaging has also become part of clinical trials, particularly in therapeutic areas such as oncology, neuroscience, and cardiovascular disease. As molecular imaging technologies have advanced beyond traditional anatomic imaging (with its emphasis on detailed views of bones, organs, and tissues), it is now possible to monitor the action of new drug candidates on the human body. *Functional imaging* has caused a shift from pure anatomic imaging to the visualization of cellular and molecular processes in living tissues. Application of biomedical and molecular imaging to the drug development process is a new technique for early identification and determination of adverse effects. Additionally, it is used for validation of efficacy, identifying which patients may respond well to the treatment, not respond at all, or be prone to a severe adverse event episode.

The need to support the acquisition, management, archival, and analysis of imaging data is similar to IT requirements in other environments, such as clinical patient care. What sets the biopharmaceutical industry apart, however, is the need to gain global regulatory approval for new medical treatments. It is therefore important to standardize measurements carried out by imaging devices and to standardize data types and interfaces. As imaging data is

integrated and incorporated into New Drug Applications (NDAs), it will be important to develop IT architectures that relate imaging data to phenotypic clinical patient data and associated genotypic data and to create applications for the query and analysis of the various data types.

In this paper, we discuss IT solutions supporting the use of imaging biomarkers in biopharmaceutical research and development. We cover clinical-trial standards created to facilitate the exchange and semantic understanding of information. We begin by discussing the current state of imaging technologies and their use in drug discovery and development. We then present a few disease-area-specific examples along with related IT requirements. Finally, we propose a high-level open-standards-based IT architecture for imaging biomarkers in biopharmaceutical research and development.

Imaging technologies and biopharmaceutical research

For nearly 70 years, medical imaging has been dominated by conventional film and screen X-ray imaging. However, during the last three decades, this field has experienced major technological growth, resulting in the development and commercialization of a plethora of new imaging technologies, introduced and briefly explained in this section. These new modalities have all been valuable additions to the clinician's arsenal of imaging tools for ever more reliable detection and diagnosis of disease.

Contrasting imaging technologies, which exploit the absorption properties of organic matter, provide the means to observe molecular entities noninvasively and nondestructively, *in vivo*, and over time. In this modality, the molecular entity being viewed is a molecular target, such as a protein in a given pathway or a small molecule that interacts with cellular processes and its environment. The application of molecular imaging enables observation of the results of a drug on a drug target, as well as its effects on a cell. This type of imaging spans the whole biopharmaceutical research and development process[11–13] and has great potential benefit.

Visualization of basic cellular processes *in vivo* provides great insights into the understanding of disease and the underlying molecular machinery. It contributes to the evaluation of drug candidates in lead optimization (i.e., the process of selecting the right drug candidate from a list of compounds) and the elucidation of efficacy, toxicology, and pharmacokinetics in preclinical studies. The nondestructive nature of molecular imaging allows for observing disease progression in live organisms. It is particularly suitable for monitoring biomarkers in living organisms. In combination with endoscopy, this technology is paving the route to new diagnostic methods and consequently, better and safer treatments.

The imaging modalities are distinguished according to their underlying physics. *Optical imaging*, the detection of photons after their interaction with tissue, basically falls into two categories, *bioluminescence imaging* (BLI) and fluorescence imaging, in particular *near infrared fluorescence* (NIRF)

■ Biomarkers are believed to hold the promise of increased productivity for biopharmaceutical research and development ■

imaging. BLI detects enzymatically generated luminescence. Luciferin-luciferase is the enzyme-substrate pair most commonly used with BLI. BLI is highly sensitive and is being applied mostly to identify qualitatively whether the luciferase reporter gene is active, indicating whether a specific pathway might be active. In fluorescent imaging, a fluorescent dye is stimulated by an external light source and emits light at a lower wavelength. Green fluorescent protein (GFP) is the dye most commonly used. Like the luciferase system, GFP can be fused to other proteins and allows high resolution imaging. Though green light does not penetrate a body very deeply, the method can be used for imaging near the surface, or in naked skin mouse models. Due to the nature of infrared light, NIRF dyes allow for imaging of structures up to 30 mm *in vivo*. Smart probes, dyes that need to be chemically activated before they show fluorescence, are used for imaging enzymatic activity, thus enabling the visualization of drug-target interaction.

Nuclear imaging, such as single-photon-emission computed tomography (SPECT) and positron emis-

sion tomography (PET), require the administration of radioactive reporter molecules. Typical applications are monitoring drug distribution, pharmacokinetics, and pharmacodynamics. As many small-molecule drugs can be labeled by using these technologies with minimal effect on the physicochemical properties, nuclear imaging has excellent potential for tracing the consequences and distribution of a chemical compound.

Magnetic resonance imaging (MRI) provides information on proton density and displays excellent contrast properties for soft tissues. MRI provides direct visualization of disease processes. For instance, in stroke models, the oxygenation deficits and subsequent membrane breakdown at later stages in the pathology can be localized precisely over the course of weeks. *Computed tomography* is well-suited to visualize bones but does not provide the best view of soft tissue; in contrast, MRI presents excellent soft tissue contrast properties. Modern approaches combine the two. Simultaneous application of paramagnetic or super-paramagnetic reporter agents allows for the simultaneous detection of molecular targets and anatomy in cancer, inflammation, and Alzheimer's disease, for example.

Ultrasound imaging is used to effectively present soft tissue (it does not apply well for imaging bones). As short pulses of sound waves at frequencies of 1 to 13 MHz are transmitted into tissue, the echoes of the waves reflect the different acoustic properties of tissues and organs and allow for the construction of an image in real time. *Ultrasound imaging* is widely used in medicine and well-known in prenatal care. This imaging modality is well-suited to the detection and visualization of moving particles, such as blood flow in vessels. By means of the Doppler effect, the velocity of the bloodstream can be quantified dynamically in the beating heart; this technique has wide applicability in the field of echocardiography.

Characteristics of imaging modalities

No single imaging technology is sufficient to cover all applications in biopharmaceutical research and development. For instance, MRI provides high spatial resolution, yet is limited with regard to sensitivity; PET and optical imaging have rather complementary features—excellent sensitivity but limited spatial resolution. Biopharmaceutical researchers have to select the imaging technologies that fit the therapeutic areas addressed by their drug discovery research.[14] For instance, CT and MRI[15–17] can be used to look at the shape of cancer tumors, whereas fluorodeoxyglucose (FDG)-PET[18] is the preferred method to analyze glucose uptake in tumors, an important measurement of tumor activity and growth. In the area of cardiovascular disease, ultrasound techniques have been applied to the study of atherosclerosis.[19]

EXAMPLES OF IMAGING BIOMARKERS

The biopharmaceutical industry is engaged in initiatives to develop biomarkers that can be used in the context of drug development.[20–22] New findings in genomics and proteomics (i.e., the study of proteins, their structures, and their function) point to various biomarkers of genetic mutation and the corresponding proteins that cause disease. In addition to conventional biomedical imaging techniques used during clinical trials, molecular imaging techniques are being developed to show how cells react in disease conditions. Imaging biomarkers may include any anatomic, physiological, biochemical, or metabolic compound that can be detected and measured with an imaging agent. In general, a biomarker must have a tight coupling to the disease process. A few disease-specific examples, described in the following subsections, illustrate this point.

Example 1: Guanylyl cyclase C as an anatomical marker and target for colorectal cancer

Guanylyl cyclase C (GCC) is a receptor protein normally found in high concentrations on the surface of the gastrointestinal epithelium. In metastatic colorectal cancer, it is present inside the cell. GCC is not expressed by tumors other than colorectal tumors. Abundant levels of GCC mRNA have been detected in human colorectal tumors and cell lines, regardless of stage and grade. Thus, GCC has potential use as a marker to determine the spread of colorectal cancer to lymph nodes. A study of 21 patients after surgical resection of colorectal cancer found that all patients who were free of cancer for five years or more (11 of the 21) were negative for GCC in lymph nodes, whereas all patients whose cancer returned within three years of surgery (the remaining 10) were positive for GCC.

GCC is a target for *in vivo* delivery of imaging agents to metastatic colon tumors. This is because STa (5–18) is a 14-amino acid peptide that selec-

tively binds to the extra-cellular domain of GCC with great affinity. STa (5–18) administered intravenously selectively recognizes and binds to GCC expressed by human colon cancer cells *in vivo*. This characterstic helps in the development of novel targeted imaging and therapeutic agents for treatment of metastatic colorectal tumors in humans.[23]

Example 2: Serum biomarkers in cardiac disease

Some of the widely known biochemical markers include Troponin, NT-proBNP (B-type natriuretic peptide), and creatine kinase. Pregnancy-associated plasma protein-A (PAPP-A) has been used as a marker for unstable plaques. Circulating markers indicating the instability of atherosclerotic plaques could have diagnostic value in unstable angina or acute myocardial infarction. The levels of PAPP-A in eight unstable coronary plaques and four stable plaques from eight patients were measured from patients who had died suddenly of cardiac problems. High levels were found in patients with unstable angina or acute myocardial infarction in contrast with levels in patients with stable angina and controls. The levels correlated with other proteins known to be involved in heart disease, namely C-reactive protein and insulin-like growth factor 1. PAPP-A is a new candidate marker for unstable angina and acute myocardial infarction.[24]

Apart from immunological detection, noninvasive methods, such as *in vivo* high-resolution MRI of atherosclerotic lesions, have been used in animal models. Cardiac imaging with echocardiography and radionuclide techniques has played an increasingly important role in cardiovascular care over the past decade.

A variety of potential cardiac imaging biomarkers are available for assessment of myocardial viability in acute and chronic ischemic heart disease. These include PET imaging for the assessment of myocardial perfusion and metabolism, SPECT imaging using Thallium 201, and dobutamine wall motion studies using echocardiography, MRI, or CT. Additional candidate approaches include contrast echocardiography, proton MRI contrast imaging and tissue tagging, Phosphorus 31 NMR spectroscopy, sodium MRI, and proton MRI to detect myocardial production of Oxygen 17 water. The latter example involves a study where magnetic resonance (MR) tagging was used to quantify the intramyocardial response to low-dose dobutamine, and to relate this response to the return of function in patients after their first myocardial infarction. The steps involved in this example are MRI, image analysis, data analysis and interpretation, and statistical analysis. It was found that there was an increase in %S (i.e., a measure of circumferential segment shortening) with peak dobutamine in dysfunctional myocardium. Dysfunctional tissue after myocardial infarction demonstrates a larger contractile response to dobutamine than normal tissue.[25]

Example 3: Neuroimaging for Parkinson's disease

Parkinson's disease is evaluated clinically if the patient presents two of three cardinal motor signs (tremor, rigidity, and bradykinesia [the slowing down and loss of spontaneous and voluntary movement]) and a response to levodopa (a drug which is highly effective in controlling most symptoms of Parkinson's disease). There are reports which suggest that 29 percent of patients initially diagnosed with Parkinson's disease by primary physicians are misdiagnosed.[26] Functional neuroimaging using SPECT provides information on the integrity of the dopaminergic system *in vivo* and thus is a useful diagnostic tool to detect early Parkinson's disease. Neuroimaging studies in association with SPECT or PET imaging identify individuals with Parkinson's disease and distinguish them from healthy subjects. A decrease in DAT (dopamine transporter) density of greater than 30 percent as compared with the healthy controls is considered to indicate neuronal degeneration and a positive diagnosis of positive Parkinson's disease.[27]

■ New findings in genomics and proteomics point to various biomarkers of genetic mutation and the corresponding proteins that cause disease ■

Example 4: Biomarkers in oncology

There are various clinical assays used routinely in the diagnosis of particular cancers that show a correlation to the presence of the tumor and enable them to be used as biomarkers for monitoring the response to cancer treatment, including serum prostrate antigen and serum CA-125 antigen (for ovarian cancer). The levels of these markers may

change due to factors not related to cancer, making correlation with tumors difficult. Combination of these markers with other markers (like those used with molecular and functional imaging) is beneficial in this regard. New imaging modalities, radioligands (i.e., radioactively labeled drugs that can associate with a receptor, transporter, enzyme, or any site of interest in the body), and contrast agents support the noninvasive visualization and quantitative measurement of physiological and molecular aspects of the tumors. The most widely used imaging technologies in oncology[28,29] are dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) and PET. For example, DCE-MRI can be used to measure tumor vascular function. Similarly, FDG-PET is used to monitor tumor metabolism before and after administration of a drug. Recently, systems that combine PET scanners and CT scanners have been introduced, enabling the detection of recurrent cervical carcinoma, for example, using PET/CT with ^{18}F-FDG (the glucose compound 18F-flurodeoxy-glucose). Imaging revealed an increase in uptake of ^{18}F-FDG. Metastasis was confirmed by biopsy.[30]

Related IT requirements

In this section, we describe the IT requirements related to the imaging technology used in Example 1. In this case, the histology lab scans the glass slides and creates digital slides, which are then reviewed by the pathologist on a computer monitor. Additionally, the slides can be analyzed with image analysis software and shared with anyone in the world (this is an example of "virtual microscopy").

There are currently no DICOM (Digital Imaging and Communication in Medicine) standards for capturing images from microscopic slides, and current IT infrastructures are challenged by image data file sizes and virtual microscopy requirements. Based on a typical glass slide size of 2.6 cm $\times$ 7.6 cm, a tissue size of 1.9 cm $\times$ 2.75 cm, and scanning at a medium power of 21,260 pixels/cm, one obtains 7 GB image files. High power gives twice the resolution in both the x and y dimensions, leading to image files of (7 GB $\times$ 2 $\times$ 2) = 28 GB. This image only represents a single plane of focus. Compression of the image can reduce the file size to about (or below) 1 GB.

In addition to the regular histological staining methods, cellular imaging systems have been developed to aid in the quantitative analysis of cellular events and the visualization of the phenotypes of the cells. For example, neurite outgrowth of the rat neuronal cell line (pheochromocytoma cells) can be detected by fluorescent staining and quantified by software. Screening of changes inside the cells is possible with the use of fluorescent-labeled antibodies. Imaging platforms with high resolution analysis and high throughput can generate about one million data points per day. Each data point is linked to the image from which it is generated. High-throughput screening technologies, integrated with analysis applications and data-storage capabilities for the images, are essential. Due to the increased interest in identifying the mode of action of drugs and in reducing adverse drug reactions, the demand for fluorescent probes in cellular imaging systems in clinical settings is increasing.

For MR and CT systems, there is a need for image acquisition and reconstruction. The MR image reconstruction task is a memory- and CPU-bound scientific computing workload. Workload requirements for CT systems today consist of processing up to 192 images per second and supporting data transfer rates of up to 300 MB/sec. Imaging data management needs can be addressed with emerging customizable content management solutions such as the IBM Content Management Offering (CMO). Other IT infrastructure needs can be addressed with server and storage products. Application software is then needed to support the analysis and visualization of the images. Therapeutic imaging often requires color and 3D versions of CT and MR images.

The following requirements have emerged for managing imaging data generated during the biopharmaceutical research and development process. An image mark-up standard must be developed; free open-source annotation, creation, and display tools, protocols for using these tools in a standardized manner on a variety of displays, and reference data sets for imaging should be made available.

A common imaging vocabulary is needed, along with a standards-based vocabulary for radiology and allied imaging fields. Natural language processing tools are needed for performing data mining in radiology reports. A set of tools is required for automatic change assessment in pixel data. Improved tools that facilitate deidentification should also be developed.

Imaging standards are needed for small animal studies, especially to support the area of digital pathology. The potential of a grid mechanism to provide functional multi-institutional and multisite services should be explored, and standards should be developed for normalized data from mammography, PET/CT, and other modalities.

FDA INITIATIVES IN IMAGING BIOMARKER-BASED CLINICAL DATA SUBMISSION

Influenced by the Critical Path Initiative of the FDA, many biopharmaceutical companies are pursuing biomarker-based clinical development initiatives aimed at safer and more efficacious drugs and improved time to market. The Division of Medical Imaging and Radiopharmaceutical Drug Products at the FDA is actively promoting a new avenue for sponsors to submit imaging biomarkers as part of the clinical submission of early drug candidates under exploratory Investigational New Drug (IND) programs, to identify promising drug candidates. The FDA promotes open-ended exploratory INDs, in which new imaging biomarkers can be introduced to help strengthen the chances of approval of a new drug candidate. It is critical that sponsors can demonstrate reproducibility and precision in their imaging findings across multisite studies and validate their results with the IRC (the Independent Image Review Charter, which reviews images collected in clinical trials for regulatory submission to ascertain the validity of findings reported from the images). The FDA mandates that archives for the submitted imaging data should be able to retain the images for possible future re-examination, and should be able to retrieve images for single and multiple trials, reanalyze images and digital data, and relate images to effective outcome assessments.

THE NEED TO STANDARDIZE IMAGING PROTOCOLS AND IT

The FDA is under considerable public pressure to optimize the review cycle of NDAs and Biologic License Applications (BLAs) so that safe and effective medications can be brought to market quickly. Every day of delay can cost biopharmaceutical companies millions of dollars in lost revenue. The expiration of drug-related patents and the emergence of strong generic drug manufacturers have prompted the biopharmaceutical industry to re-engineer its research and development processes and to look for ways to use technology to cut costs and speed up development.

To ascertain the risks involving safety and efficacy of new drug candidates, the FDA has determined that it needs tools to compare data on new drugs to data on other drugs in the same therapeutic area and drug class. Therefore, to enable efficient review of electronic clinical data submissions and to support cross-trial analysis, the FDA has recommended the Study Data Tabulation Model (SDTM) of the Clinical Data Interchange Standards Consortium (CDISC) as the standard for drug submissions, specifically the use of the SDTM 3.1 format for submission of clinical study data tabulations in the Study Data Specification guide.[31] The FDA has spent considerable time working with CDISC representatives, giving input and direction during the development of the SDTM. Traditionally, most drug applications included traditional clinical endpoints, but based on recent submission activities, it is evident that use of biomarker data as surrogate endpoints is becoming a valid alternative. The SDTM standards support submission of standardized data for both traditional laboratory test-based findings as well as the emerging genomic-based and imaging biomarker-based results.

CDISC SDTM is an easily extensible model that incorporates the data structures necessary to capture the submission data to be sent to the FDA. It gives the FDA a standard format for all clinical trial submissions. Because the standard was developed with strong collaboration between the biopharmaceutical industry, clinical research organizations, clinical trial sites, IT vendors, and the FDA, it represents the collective input of a broad group of stakeholders.

Table 1 shows four major data categorizations or classes of the SDTM data model. These categorizations were designed to simplify the model. The "other" class is reserved for specialized areas. The "related records" domain in this class is a mechanism to provide linkages across the different files (i.e., domains) within a class or across multiple classes.

As of September 2006, two new SDTM domains have been designed to support biomarker data submission. The pharmacogenomics (PG) and pharmacogenomics results (PR) domains will support the submission of summarized genomic data. Efforts will be underway soon to collect sample data from the industry in order to validate the PG

Table 1 Data classes of the SDTM data model

Interventions	Events	Findings	Other
Concomitant medications	Adverse events	Questions	Trial design
Exposure	Dispositions	Electrocardiograms	Related records
Substance use	Medical histories	Laboratory results	Supplemental qualifiers
		Physical examination results	Trial summaries
		Vital signs	
		Subjective characterizations	
		Inclusions/Exclusions	

and PR domains. It is expected that additional changes may evolve from that effort. In addition, an imaging (IM) domain is being proposed that will include a mapping of the relevant DICOM metadata fields required to summarize an imaging data submission.

Figure 1 shows the PG and PR domains, which are part of the findings class and are designed to store pharmacogenomics panel ordering information. The detailed test-level information, such as genotype/SNP (single nucleotide polymorphism) summarized results, are reported in the PR domain. The example in the figure shows what a typical genotype test might look like in terms of data content and usage of the HUGO (Human Genome Organization) nomenclature.[32] The PG domain supports the hierarchical nature of pharmacogenomic results, where for a given genetic test from a patient sample (listed in the parent domain), multiple genotypes or SNPs can be reported (and listed in the child domain).

A sample mapping of DICOM metadata tags into the fields of the IM domain is shown in *Table 2*. The designs of the new PG and IM domains are currently being vetted among the various CDISC and FDA stakeholders as a step toward their finalization.

Although the FDA has proposed the SDTM data model for submission data, this is only an interchange format for sponsors to submit summarized clinical study data in a standardized fashion to the FDA. The FDA has also identified a need for a relational repository model to store the SDTM data sets. The requirement was to design a normalized and extensible relational repository model that would scale up to a huge collection of studies going back into the past and supporting those in the future. Under a Cooperative Research and Development Agreement (CRADA), the FDA and IBM have jointly developed this repository for submissions, the JANUS model (named after the two-headed Roman god) that can look backward to support historic retrospective trials and forward to support prospective trials. JANUS refers both to the open-source data model and the repository that implements that model. As shown in *Figure 2*, the data classification system of CDISC with classes such as interventions, findings, and events was leveraged in the JANUS model with linkages to the subjects (for the patients enrolled in the clinical trial) to facilitate navigation across different tables by consolidating data in three major tables. Benefits resulting from this technique include reduced database maintenance and a simpler data structure that is easier to understand and can support cross-trial analysis scenarios. The ETL (Extract-Transform-Load) process for loading the SDTM domain data sets instantiates the appropriate class table structure in JANUS without requiring any structural changes.

DATA INTEGRATION CHALLENGES

There are a number of challenges associated with the integration of clinical and biomarker data. These include the lack of standardized vocabulary definitions throughout the industry and changing business definitions for the core elements, which cause a divergent set of views throughout the industry. External sources that bring in source data, such as PACS (Picture Archiving and Communications Systems) systems for imaging, ArrayTrack for genomic data submission, external reference databases such as PubMed,[33] GenBank,[34] dbSNP,[35]

Parent Domain:

STUDYID	USUBJID	PGSEQ	PGGRPID	PGREFID	PGOBJ	PGTESTCD	PGTEST
NSCLC10	ZB1000-0007	1	EGFR-KD-001	SPEC001		EGFR-KD	EGFR-KD (EGFR Gene, Protein)
NSCLC10	ZB1000-007	7	CYP1A2-00001	SPEC002		CYP1A2	CYP1A2 Mutation
NSCLC10	ZB1000-008	1	CYP1A2-00003	SPEC001		CYP1A2	CYP1A2 Mutation DNA Analysis
NSCLC10	ZB1000-009	1	CYP2D6-00001	SPEC001		CYP2D6	CYP2D6 test
NSCLC10	ZB1000-009	11	CYP2C19-00001	SPEC001		CYP2C19 (*2,	Cytochrome P450 2C19 Test

PGMETHCD	PGASSAY	PGORRES	PGSTRESC	PGSTRESN
88323, 88380, 83890 (X2), 838	12700056	EGFR	EGFR	
83891, 83892 x2, 83998	50-776	CYP 1A2	CYP 1A2	
83891, 83892 x2, 83998	50-777	CYP1A2	CYP1A2	
83891, 83892, 83901 x2,	50-574	CYP2D6	CYP2D6	
83891, 83892, 83901 x2,	50-575	CYP2C19	CYP2C19	

Child Domain:

STUDYID	USUBJID	PGSEQ	PGGRPID	PGREFID	PGOBJ	PGTESTCD	PGTEST
NSCLC10	ZB1000-009	1	CYP2D6-00001	SPEC001	HGNC:2625	CYP2D6	CYP2D6 GENE.g.-1584C>G
NSCLC10	ZB1000-009	2	CYP2D6-00001	SPEC001	HGNC:2625	CYP2D6	CYP2D6 GENE.g.100C>T
NSCLC10	ZB1000-009	3	CYP2D6-00001	SPEC001	HGNC:2625	CYP2D6	CYP2D6 GENE.g.124G>A
NSCLC10	ZB1000-009	4	CYP2D6-00001	SPEC001	HGNC:2625	CYP2D6	CYP2D6 GENE.g.883G>C
NSCLC10	ZB1000-009	5	CYP2D6-00001	SPEC001	HGNC:2625	CYP2D6	CYP2D6 GENE.g.1023C>T
…	…	…	…	…	…	…	…

PGMETHCD	PGASSAY	PGORRES	PGSTRESC	PGSTRESN
MOLGEN	50-574	M33388:g.-1584GG		
MOLGEN	50-574	M33388:g.100TG		
MOLGEN	50-574	M33388:g.124GC		
MOLGEN	50-574	M33388:g.883GC		
MOLGEN	50-574	M33388:g.1023CG		
…	…	…		

Figure 1
Partial sample of pharmacogenomics SDTM domains

SwissProt,[36] and others are not integrated. There is no consensus on what parts of genomic data elements are crucial for understanding clinical outcomes. Genomics does not fit simply into the clinical assessment model. Imaging data from various clinical sites is heterogeneous in nature, but a uniform and standardized review environment is required for independent reviewers in imaging Contract Research Organizations (CROs) to annotate and mark up the images and to substantiate a study's hypothesis through analysis of the findings. These challenges are discussed in more detail in the following subsections.

Standardization of vocabulary definitions

Vocabulary definitions have not been standardized, and laboratories tend to use their own codes to identify genomic tests. There are Logical Observation Identifier Names and Codes (LOINC**) codes for some disease gene mutations, and new LOINC codes need to be developed for other gene mutations. The use of standardized vocabularies or terminologies is required in order to fully exploit the cross-trial capabilities of the JANUS repository. They are critical to establishing a common understanding of clinical data that supports consistent analysis.

Because genomics is a relatively new field in research, different organizations use and define data within various contexts. As the science behind genomics is better understood, business definitions are modified to better represent these new discoveries. As a result, there are discrepancies in the business definitions of different organizations.

Integration of data sources

Integration of data sources such as GenBank, SwissProt, and dbSNP can be complicated, especially if

Table 2 Partial mapping of DICOM imaging metadata tags to SDTM IM domain fields

CDISC SDTM IM Domain		DICOM Tags		
Variable Name	CDISC Notes (for domains) or Description (for general classes)	Tag	Attribute Name	Attribute Description
Unique subject identifier	Unique subject identifier within submission.	(0012, 0040)	Clinical trial subject ID	The assigned identifier for the clinical test subject; shall be present if clinical trial subject reading ID is absent; may be present otherwise.
Sequence number	Sequence number given to ensure uniqueness within a data set for a subject. It can be used to join related records.	(0020, 0013)	Instance number	A number that identifies this image. **Note**: this attribute was named Image Number in earlier versions of this standard.
Imaging reference ID	Internal or external identifier. Example: UUID for external imaging data file.	(0008, 0018)	Standard operating procedure instance UID	Uniquely identifies the standard operating procedure instance.
Test or examination short name	Short name of the measurement, test or examination. It can be used as a column name when converting to a data set from a vertical to a horizontal format.	(0008, 1030)	Study description	Institution-generated description or classification of the study (component) performed.

their use by evolving systems does not match actual laboratory use. Standardized vocabularies (i.e., ontologies) will link these data sources for validation and analysis purposes. These data sources tend to represent the frontiers of science, especially because they store genetic biomarkers associated with diseases and best methods of testing which are continually evolving. Having a reliable link between genetic testing laboratories, external data sources for innovations in medical science, and clinical data greatly improves analytical functionality, resulting in more accurate outcome analysis. These links have been designed into the CDISC PG and PR domains to facilitate the analysis and reporting of genetic factors in clinical trial outcomes.

Consensus on significance of genomic data

Another obstacle commonly encountered is the lack of consensus on what genetic attributes are crucial to the analysis of clinical outcomes. This is an evolving area and therefore likely to change. However, careful use of ontologies may at least provide a way of normalizing a core set of data elements that could be used in cross-study analysis. Much of the information that is textual in nature needs a stronger method of categorization so that subjective analysis, which tends to be categorical in nature, can have consistent definitions.

Semantic interoperability

As standards have continued to evolve, the need for semantic interoperability has become quite clear. In order to effectively use standards to exchange information, there must be an agreed-upon data structure, and the stakeholders must share a common definition for the data content itself. The true benefit of standards is attained when two different groups can reach the same conclusions based on access to the same data because there is a shared understanding of the meaning of the data and the context in which it is used. Standards must cover a wide variety of stakeholders within the health-care and life-science industries. The development of business definitions within a metadata repository is indispensable whether one wishes simply to share information within an organization or across a large spectrum of stakeholders that might include pharmaceutical companies, clinical research organiza-

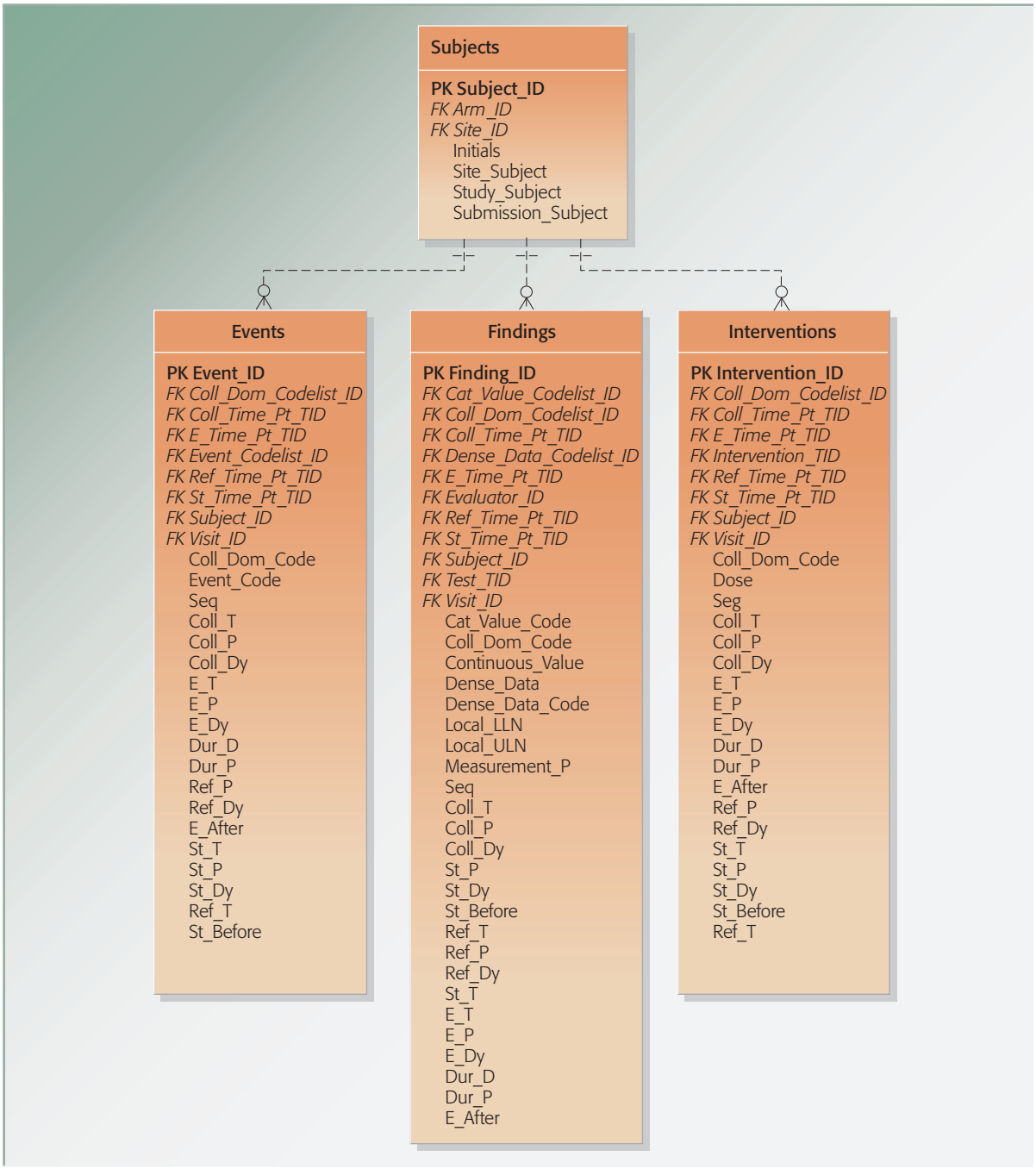

Figure 2
Core entities in the JANUS data model

tions, laboratories, medical research centers, healthcare providers, public-health agencies, and clinical regulatory agencies.

The FDA requires imaging findings to be reproducible so that an independent reviewer can draw the same conclusion or derive the same computed measurements as those included in a submission. As a result, a unified architecture is required for a DICOM-based imaging data-management platform that supports heterogeneous image capture environments and modalities and allows Web-based access to the independent reviewers. Automated markups and computations are recommended to

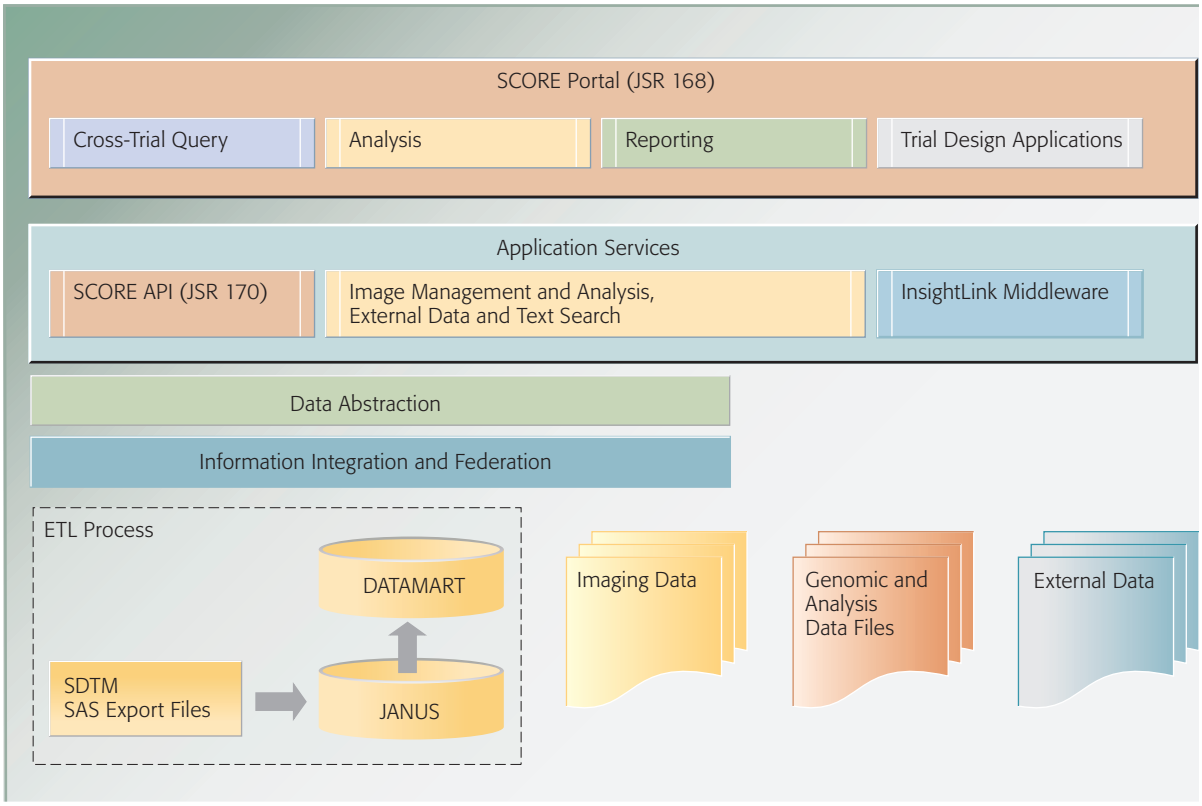

Figure 3
Proposed reference architecture for biomarker-based clinical development

facilitate reproducibility, but manual segmentation or annotations are often needed to compute the imaging findings. A common vocabulary is also needed for the radiological reports that specify diagnosis and other detailed findings and for the specification of the imaging protocols.

PROPOSED IT ARCHITECTURE FOR IMAGING BIOMARKERS

Based on the technical challenges and requirements inherent in integrating a diverse set of data sources for a biomarker-based clinical data submission, we propose a reference architecture (shown in *Figure 3*) that addresses a majority of those requirements. Although this architecture includes software products and assets designed by IBM, it can logically be extended to fit other vendors' products as well. Our approach is to present a general-purpose platform for managing clinical submissions of imaging biomarker data, in contrast to the specialized portals proposed by Pivovarov et al.[37] and Amies et al.[38]

At the lowest data layer, summarized clinical submission data in the SDTM format feeds (as exported by SAS) into JANUS from a Clinical Data Management System (CDMS)[39,40] that stores CRFs (Case Report Forms). The associated metadata for the SDTM submission is mapped into the tables in the JANUS repository. Because JANUS is a normalized repository format optimized for efficient storage (using partitioned indexes), one needs to build a collection of application- and use-case specific *datamarts* (i.e., relational data models created on top of data stores or data warehouses for supporting more efficient and faster querying) on top of JANUS. Aside from the core submission data in JANUS, one would need to link with the imaging data in the PACS systems that can be centrally managed with a standardized imaging broker service, such as that provided by CMO, with the genomic raw and analysis files stored in ArrayTrack and the content management repository supported by SCORE (Solution for Compliance in a Regulated Environment),[41] and finally, with external reference databases such as PubMed, GenBank, dbSNP, and

SwissProt. The external reference would be linked by using unstructured information management technology provided, for example, by WebSphere* Information Integrator or OmniFind*. All of these content repositories can be searched dynamically by using a federated warehouse constructed by Information Integrator,[42] which uses a wrapper-based technology for linking diverse data sources.

On top of the federation layer, we propose a data abstraction layer powered by Data Discovery Query Builder (DDQB),[43] which exposes a user-centric logical data model (based on XML [Extensible Markup Language]) that is mapped on top of the physical data model. DDQB is a technology component developed for the Mayo Clinic and deployed in a number of biobank and clinical genomics projects.

In the application services layer, we propose a JSR-170-compliant[44] API (application programming interface) for analytical applications to store their results into the JCR (Java** Content Repository) managed by SCORE. The imaging data is available for quick viewing in thin Web clients (e.g., browsers) next to the clinical outcome data by using a servlet architecture proposed by an emerging DICOM standard called WADO (Web Access for DICOM Objects).

For collaboration at this layer, we present the innovative InsightLink solution that is linked with data entities mapped to the semantic Web by unique URI-type (Uniform Resource Identifier type) identifiers called Life Sciences Identifiers (LSIDs). InsightLink is a service-oriented-architecture (SOA)-based middleware that provides a flexible platform for managing a variety of annotation types (using predefined XML forms) mapped on top of a variety of data formats (PDF, Microsoft Office, Web pages, and relational data elements). There is flexible API support (for COM [Common Object Model], SOAP [Simple Object Access Protocol], PERL [Practical Extraction and Reporting Language], and native Java) provided so that applications can integrate annotation functionality within their existing interfaces using a plug-in architecture.

Finally, we propose an integrated portal-based collaborative environment based on SCORE for launching clinical data querying and analysis tools within a 21CFRPart11-compliant environment (211CFRPart11 is a set of FDA compliance regulations for electronic records and signatures in the biopharmaceutical industry). The JSR-168[45] open standard for portlets supports interoperability of portlets between portal technologies of multiple vendors. In addition to the collaboration platform promoted by SCORE, it also allows a business-choreography-based workflow design and execution framework for integrating business processes, such as markup and annotation of images for computing surrogate endpoints from the images included in the CRF, after independent review for quality assurance.

CONCLUSIONS

Aside from the development of faster, more inexpensive computing capabilities, significant advances have been made in the signal and image-processing theories on which the development and maturation of many new imaging technologies are based. In addition, the rapid development and deployment of methods for archiving and transmitting digital images have allowed hospitals to distribute an increasing number of images and associated diagnoses in a timely and cost-effective fashion.

Although still undergoing significant advances toward higher sensitivity and specificity, improved resolution, and image quality, medical imaging in clinical care has made significant advances. It is a maturing field with data management needs that are quite well understood and served by conventional PACS systems. Imaging data management requirements in biomedical research and biopharmaceutical research and development are quite different from those in clinical care. High-throughput imaging of cell structure and protein localization and its relation to other data sets (e.g., microarrays) at the systems biology level is rapidly expanding, leading to data expansion and subsequent IT challenges. Because the goal of biopharmaceutical companies is to discover and develop medical treatments in a regulated environment, biomedical and molecular imaging procedures must be standardized, measurements must give reproducible results even in multicenter clinical studies, and the associated data must be managed with great care.

Though small compared with the medical imaging market in health care, the biopharmaceutical imaging market is highly important and strategic. Healthcare providers will eventually have to adopt standards for the validation and measurement of imaging biomarkers that will be agreed upon by the

industry in cooperation with the medical research community, medical device manufacturers, and the FDA. In addition, clinical care providers will eventually adopt the new diagnostic procedures and medical treatments enabled by the use of advanced imaging biomarkers.

In this paper, we have described ongoing efforts by the industry to translate ideas like the FDA's Critical Path Initiative into tangible improvements of the research and development process. By using imaging biomarkers in therapeutic areas such as oncology, neuroscience, and cardiovascular disease, biopharmaceutical companies are taking advantage of new imaging technologies to develop safer and more efficacious medical treatments, and to shorten lead times in bringing these treatments to patients.

Significant new initiatives such as the FDG-PET Lymphoma Project[46] co-sponsored by NCI, the FDA and CMS (the Centers for Medicare and Medicaid Services) and emerging standardization efforts by NIST (National Institute of Standards and Technology) are indicators of progress in this area. NCI's RIDER (Reference Image Database to Evaluate Response to Drug Therapy in Lung Cancer) project is another specific step in this direction. The Alzheimer Disease Neuroimaging Initiative (ADNI)[47] is an initiative in neuroscience to test whether serial MRI, PET, biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease.[48]

As CROs add imaging data management capabilities (or outsource those activities to imaging core labs), the industry is encouraged to incorporate imaging data in New Drug Applications. However, significant IT challenges have to be addressed before such applications become routine and are dealt with effectively by both the industry and the FDA.

It is our opinion that the way forward is to adopt open standards such as SDTM and extensions of JANUS, and to adopt robust and scalable IT architectures, such as those outlined in this paper. IBM middleware products or compatible alternatives are proposed as the solid backbones of such architectures. Solutions such as SCORE and CMO can be customized and combined to satisfy the requirements of image data management in a biomarker-based biopharmaceutical research and development environment.

*Trademark, service mark, or registered trademark of International Business Machines Corporation in the United States, other countries, or both.

**Trademark, service mark, or registered trademark of Regenstrief Foundation, Inc. or Sun Microsystems, Inc.

CITED REFERENCES

1. J. A. DiMasi, "The Value of Improving the Productivity of the Drug Development Process: Faster Times and Better Decisions," *PharmacoEconomics* **20**, Supplement 3, 1–10 (2002).
2. J. A. DiMasi, R. W. Hansen, and H. G. Grabowski, "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics* **22**, No. 2, 151–185 (2003).
3. *Pharma 2010: The Threshold of Innovation*, IBM Business Consulting Services, http://www.ibm.com/industries/healthcare/doc/content/resource/insight/941673105.html?g_type=rhc.
4. F. W. Frueh, *PGx and BIOMARKERS: Current Role in Drug Development*, U. S. Department of Health and Human Services, Food and Drug Administration (June 2005), http://www.fda.gov/cder/genomics/PGX_biomarkers.pdf.
5. *Challenge and Opportunity on the Critical Path to New Medical Products*, U. S. Department of Health and Human Services, Food and Drug Administration (March 2004), http://www.fda.gov/oc/initiatives/criticalpath/whitepaper.html.
6. L. J. Lesko and A. J. Atkinson, Jr., "Use of Biomarkers and Surrogate Endpoints in Drug Development and Regulatory Decision Making: Criteria, Validation, Strategies," *Annual Review of Pharmacology and Toxicology* **41**, 347–366 (2001).
7. J. G. Hardy and C. G. Wilson, "Radionuclide Imaging in Pharmaceutical, Physiological and Pharmacological Research," *Clinical Physics and Physiological Measurement* **2**, No. 2, 71–121 (May 1981).
8. K. Sikora, H. Smedley, and P. Thorpe, "Tumour Imaging and Drug Targeting," *British Medical Bulletin* **40**, No. 3, 233–239 (July 1984).
9. A. J. Fischman, N. M. Alpert, and R. H. Rubin, "Pharmacokinetic Imaging: A Noninvasive Method for Determining Drug Distribution and Action," *Clinical Pharmacokinetics* **41**, No. 8, 581–602 (2002).
10. J. J. Smith, A. G. Sorensen, and J. H. Thrall, "Biomarkers in Imaging: Realizing Radiology's Future," *Radiology* **227**, No. 3, 633–638 (2003).
11. M. Rudin and R. Weissleder, "Molecular Imaging in Drug Discovery and Development," *Nature Reviews Drug Discovery* Volume 2, Part 2, 123–131 (2003).
12. F. Jaffer and R. Weissleder, "Molecular Imaging in the Clinical Arena," *Journal of the American Medical Association* **293**, No. 7, 855–862 (2005).
13. H. Mucke, *Molecular Imaging Comes of Age, Applications and Impacts in Discovery, Clinical Trials, and Medical Practice*, Cambridge Health Associates (CHA) Life Sciences Reports (October 2004).

14. H. H. Pien, A. J. Fischman, J. H. Thrall, and A. G. Sorensen, "Using Imaging Biomarkers to Accelerate Drug Development and Clinical Trials," *Drug Discovery Today* **10**, No. 4, 259–266 (February 2005).
15. N. Beckmann, D. Laurent, B. Tigani, R. Panizzutti, and M. Rudin, "Magnetic Resonance Imaging in Drug Discovery: Lessons from Disease Areas," *Drug Discovery Today* **9**, No. 1, 35–42 (January 2004).
16. A. G. Sorensen, "Magnetic Resonance as a Cancer Imaging Biomarker," *Journal of Clinical Oncology* **24**, No. 20, 3274–3281 (July 2006).
17. N. Hylton, "Dynamic Contrast-Enhanced Magnetic Resonance Imaging as an Imaging Biomarker," *Journal of Clinical Oncology* **24**, No. 20, 3293–3298 (July 2006).
18. W. A. Weber, "Positron Emission Tomography as an Imaging Biomarker," *Journal of Clinical Oncology* **24**, No. 20, 3282–92 (July 2006).
19. J. T. Salonen, K. Nyyssonen, R. Salonen, E. Porkkala-Sarataho, T.-P. Tuomainen, U. Diczfalusy, and I. Bjorkhem, "Lipoprotein Oxidation and Progression of Carotid Atherosclerosis," *Circulation* **95**, No. 4, 840–845 (1997).
20. D. A. Lewin and M. P. Weiner, "Molecular Biomarkers in Drug Development," *Drug Discovery Today* **9**, No. 22, 976–983 (November 2004).
21. N. Davies, T. Peakman, and S. Arlington, "A New Formula for Finding Drugs," *Drug Discovery Today* **9**, No. 5, 197–199 (March 2004).
22. R. Frank and R. Hargreaves, "Clinical Biomarkers in Drug Discovery and Development," *Nature Reviews Drug Discovery* Volume 2, Part 7, 566–580 (July 2003).
23. A. R. Wolfe, M. Mendizabal, E. Leong, A. Cuthbertson, V. Desai, S. Pullan, D. K. Fujii, M. Morrison, R. Pither, and S. A. Waldman, "In Vivo Imaging of Human Colon Cancer Xenografts in Immunodeficient Mice Using a Guanylyl Cyclase C–Specific Ligand," *Journal of Nuclear Medicine* **43**, No. 3, 392–399 (2002).
24. A. Bayes-Genis, C. A. Conover, M. T. Overgaard, K. R. Bailey, M. Christiansen, D. R. Holmes, Jr., R. Virmani, C. Oxvig, and R. S. Schwartz, "Pregnancy-Associated Plasma Protein A as a Marker of Acute Coronary Syndromes," *New England Journal of Medicine* **345**, No. 14, 1022–1029 (October 2001).
25. G. Geskin, C. M. Kramer, W. J. Rogers, T. M. Theobald, D. Pakstis, Y. L. Hu, and N. Reichek, "Quantitative Assessment of Myocardial Viability after Infarction by Dobutamine Magnetic Resonance Tagging," *Circulation* **98**, No. 3, 217–223 (July 1998).
26. S. G. Reich, M. B. Lederman, and M. E. Griswold, "Errors and Delays in Diagnosing Parkinson's Disease," *Annals of Neurology* **52**, No. S3, p. S84 (2002).
27. D. L. Jennings, J. P. Seibyl, D. Oakes, S. Eberly, J. Murphy, and K. Marek, "(123I) beta-CIT and Single-Photon Emission Computed Tomographic Imaging vs Clinical Evaluation in Parkinsonian Syndrome: Unmasking an Early Diagnosis," *Archives of Neurology* **61**, No. 8, 1224–1229 (August 2004).
28. B. J. Hillman, "Introduction to the Special Issue on Medical Imaging in Oncology," *Journal of Clinical Oncology* **24**, No. 20, 3223–3224 (July 2006).
29. M. Atri, "New Technologies and Directed Agents for Applications of Cancer Imaging," *Journal of Clinical Oncology* **24**, No. 20, 3299–3308 (July 2006).
30. M. E. Juweid and B. D. Cheson, "Positron-Emission Tomography and Assessment of Cancer Therapy," *New England Journal of Medicine* **354**, No. 5, 496–507 (February 2006).
31. *Study Data Specifications Version 1.3*, U. S. Food and Drug Administration (2006), http://www.fda.gov/cder/regulatory/ersr/Studydata-v1.3.pdf.
32. HUGO Gene Nomenclature Committee, http://www.gene.ucl.ac.uk/nomenclature/.
33. Entrez PubMed, U. S. National Library of Medicine, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed.
34. GenBank Overview, National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/Genbank.
35. Single Nucleotide Polymorphism, National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/SNP.
36. UniProtKB/Swiss-Prot, European Bioinformatics Institute, http://www.ebi.ac.uk/swissprot/.
37. M. Pivovarov, G. Bhandary, U. Mahmood, G. Zahlmann, M. Naraghi, and R. Weissleder, "MIPortal: A High Capacity Server for Molecular Imaging Research," *Molecular Imaging* **4**, No. 4, 425–431 (October-December 2005).
38. C. Amies, A. Bani-Hashemi, J. C. Celi, G. Grousset, F. Ghelmansarai, D. Hristov, D. Lane, M. Mitschke, A. Singh, H. Shukla, J. Stein, and M. Wofford, "A Multi-Platform Approach to Image Guided Radiation Therapy (IGRT)," *Medical Dosimetry* **31**, No. 1, 12–19 (Spring 2006).
39. *Oracle Clinical*, Oracle Corporation, http://www.oracle.com/industries/life_sciences/oc_data_sheet_81202.pdf.
40. CDMS with Clintrial, Phase Forward, http://www.phaseforward.com/products_cdms_clintrial.html.
41. IBM SCORE: The Next Generation in Compliance Solutions, IBM Corporation, http://www.ibm.com/industries/healthcare/doc/content/landingdtw/1160437105.html?g_type=pspot.
42. IBM WebSphere Information Integrator, http://www.ibm.com/industries/healthcare/doc/content/solution/939513105.html.
43. IBM Data Discovery and Query Builder, http://publib.boulder.ibm.com/infocenter/eserver/v1r2/index.jsp?topic=/ddqb/eicavkickoff.htm.
44. JSR 170: Content Repository for Java Technology API, Java Community Process, http://jcp.org/en/jsr/detail?id=170.
45. JSR 168: Portlet Specification, Java Community Process, http://jcp.org/en/jsr/detail?id=168.
46. G. J. Kelloff, J. M. Hoffman, B. Johnson, H. I. Scher, B. A. Siegel, E. Y. Cheng, B. D. Cheson, J. O'Shaughnessy, K. Z. Guyton, D. A. Mankoff, L. Shankar, S. M. Larson, C. C. Sigman, R. L. Schilsky, and D. C. Sullivan, "Progress and Promise of FDG-PET Imaging for Cancer Patient Management and Oncologic Drug Development," *Clinical Cancer Research* **11**, No. 8, 2785–2808 (April 2005).
47. Alzheimer's Disease Neuroimaging Initiative Questions and Answers, U. S. National Institutes of Health, National Institute on Aging (October 13, 2004), http://www.nia.nih.gov/NewsAndEvents/PressReleases/PR20041013ADNI.htm.
48. T. Sunderland, H. Hampel, M. Takeda, K. T. Putnam, and R. M. Cohen, "Biomarkers in the Diagnosis of Alzheimer's Disease: Are We Ready?" *Journal of Geriatric Psychiatry and Neurology* **19**, No. 3, 172–179 (September 2006).

Accepted for publication August 28, 2006.
Published online December 31, 2006.

Michael Hehenberger
IBM Healthcare and Life Sciences, Route 100, Building 3, Room 1J21, Somers, New York 10589 (hehenbem@us.ibm.com). Dr. Hehenberger leads the development and implementation of IBM's global life sciences and biopharmaceutical solutions for the "Life Sciences Transformation" business segment. As part of the IBM Life Sciences initiative, he initially worked on data and knowledge management solutions for biopharmaceutical research and development. More recently, he has specialized in the integration and analysis of information related to chemical, biological, and clinical (patient) databases and on biomedical informatics, in particular clinical genomics, pharmacogenomics, and biomedical and molecular imaging. His current focus is on a broad range of life sciences solutions, including information systems supporting biomarker-based development of new medical treatments. Dr. Hehenberger holds advanced degrees in physics from the Technical University of Vienna, Austria and Ph.D. and Dr.Sc. degrees in quantum chemistry from Uppsala University in Sweden.

Avijit Chatterjee
IBM Healthcare and Life Sciences, Route 100, Building 1, Room 1J24, Somers, New York 10589 (achatter@us.ibm.com). Dr. Chatterjee is a Senior Technical Staff Member in the IBM HealthCare and Life Sciences Solutions Development organization. He is currently the lead architect of the information-based medicine solution for pharmacogenomics. Before joining the Life Sciences organization in 2001, he worked in IBM Global Services as a certified consultant, spending over seven years as a data-mining consultant. He has also worked as a lead architect, designing portals and other text-mining solutions. Dr. Chatterjee has a broad background in knowledge management and business intelligence, with expertise in data and text mining as well as visualization. He played an instrumental role in leading the launch of two IBM software products for which he has filed six patents; he has also won an Outstanding Technical Achievement award for his leadership on the collaboration technology, InsightLink. He has a Ph.D. degree in computer science from the University of Southern California.

Usha Reddy
IBM Software Services, Health Care and Life Sciences Industry Solutions, 89 Schreiner Drive, North Wales, Pennsylvania 19454 (usharedy@us.ibm.com). Dr. Reddy is a Senior Scientist working with the Healthcare and Life Sciences Industry Solutions team. Her focus for the past five years has been in the area of health care, clinical genomics, and life sciences discovery. She has played a leading role in various genomic projects within the medical research industry. Before this, Dr. Reddy spent several years in academia as a postdoctoral fellow and research faculty member at the University of Pennsylvania. Her major areas of expertise are molecular biology, biochemistry, neurobiology, and bioinformatics. She has a Ph.D. degree in biochemistry from the University of Hyderabad, India, and an M.S degree in bioinformatics from the University of Pennsylvania.

Joyce Hernandez
Merck and Company, Incorporated, 126 Lincoln Avenue, Mail stop: RY84-24, Rahway, New Jersey 07065-0900 (joyce_hernandez@merck.com). Dr. Hernandez is a senior architect specializing in providing clients with business intelligence solutions, aligning the business initiatives of the enterprise with the data warehouse infrastructure through a combination of decision support information and technology. She is widely recognized for her skills in business discovery, decision support, and data design, with a focus on performance measurements and organizational factors that are essential to an effective solution. She has implemented solutions in the health care, social services, manufacturing, and transportation industries. Her most recent projects have involved work with the FDA and the pharmaceutical industry in the development of clinical trial standards to position the regulatory agency's ability to develop a data warehouse. Dr. Hernandez has also worked with the FDA to develop a scalable data warehouse design to store current submission data and meet future needs.

Jörg Sprengel
IBM Sales and Distribution, IBM Switzerland, Schwarzwaldallee 215, CH-4010 Basel, Switzerland (j.sprengel@ch.ibm.com). Dr. Sprengel is the global research and development client manager for the integrated account of a large pharma company. Over the past 10 years, he has held different positions in the pharmaceutical and biotechnological industry, where he faced the challenging task of bridging the chasm between IT, science, and the business of drug discovery. He joined IBM in 2004. Recently, Dr. Sprengel has been involved with imaging technologies, their application, and their integration into the research and development process. He received a Ph.D. degree from the University of Cologne, Germany for performing the computational analysis of the human adenovirus type 12 genome. ■