Informatics, Data Science, and Artificial Intelligence: Building the Educational Pipeline for the Future of Radiation Oncology

Dave Fuller, MD, PhD
cdfuller@mdanderson.org
@cd_fuller
**2016-19 Funders:**

The Andrew Sabin Family Fellowship Program, through an endowment established by the Andrew Sabin Family Foundation

A direct gift from the Beach Family of Phoenix, AZ.

NIH Big Data to Knowledge (BD2K) Program of the National Cancer Institute (NCI) Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Award (1R01CA214825-01)

National Science Foundation, Division of Mathematical Sciences, Quantitative Approaches to Biomedical Big Data (QuBBD)/Big Data to Knowledge (BD2K) Program (NSF1557559);

National Cancer Institute Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science (1 R01 CA214825-01)

• National Institute of Dental and Craniofacial Research (NR56/R01 DE025248-01)

• National Cancer Institute Grant MD Anderson Head and Neck Specialized Programs of Research Excellence (SPORE) Development Award (P50CA097007-10)

• Elekta AB/MD Anderson MRI-LinAc Consortium Seed Grant*

• Elekta AB Travel support & Honoraria*

• Philantropic individuals/agencies/societies

• Corporate/industry funders

• Federal or state funding agencies
2014 Statistical and Applied Mathematical Sciences Institute (SAMSI) Innovation Lab
"Interdisciplinary Approaches to Biomedical Data Science Challenges"
Team

Liz Marai
PhD
Computer Science
U. Illinois-Chicago

David Vock
PhD
Biostatistics
U. Minnesota

Guadalupe Canahuate
PhD
Computer Science
U. Iowa

Dave Fuller
MD, PhD
Radiation Oncology
UT MD Anderson

-2016 Joint NIH/NSF Division of Mathematical Sciences Initiative on Quantitative Approaches to Biomedical Big Data (QuBBD) Grant, “Spatial-nonspatial Multidimensional Adaptive Radiotherapy Treatment” (NSF 1557679)
-2017-2020 Early Stage Development of Technologies in Biomedical Computing, Informatics, and Big Data Science Grant, “SMART-ACT: Spatial Methodologic Approaches for Risk Assessment and Therapeutic Adaptation in Cancer Treatment” (R01 CA214825-01)
Informatics, Data Science, and Artificial Intelligence

Figure 2. Traditional Research Model and Data-Driven Models in Artificial Intelligence

- Traditional research model
- Data-driven research models

Figure 1. Biomedical Research and Informatics Approaches in Artificial Intelligence
As a specialty, radiation oncology has a long track record of deploying novel technology for the benefit of our patients. AI will be a part of our future. Even though this may seem to be on the distant horizon, the opportunity cost of ignoring AI at this juncture is steep.
Artificial intelligence in radiation oncology: A specialty-wide disruptive transformation?

Reid F. Thompson\textsuperscript{a,b,*}, Gilmer Valdes\textsuperscript{c}, Clifton D. Fuller\textsuperscript{d}, Colin M. Carpenter\textsuperscript{e}, Olivier Morin\textsuperscript{c}, Sanjay Aneja\textsuperscript{f}, William D. Lindsay\textsuperscript{g}, Hugo J.W.L. Aerts\textsuperscript{h,i}, Barbara Agrimson\textsuperscript{a}, Curtiland Deville Jr.\textsuperscript{j}, Seth A. Rosenthal\textsuperscript{k}, James B. Yu\textsuperscript{f}, Charles R. Thomas Jr\textsuperscript{a}

From vision to reality: How do we get from here to there?

Recognition of the magnitude of the task, the need for collaboration with other disciplines and specialties, and the potential revolutionary, paradigm-changing nature of the changes to come in the coming years should all serve as a “call to arms” for Radiation Oncology to actively engage in harnessing the potential advantages of AI for our specialty and for our patients. The question remains “How?”
Informatics [The Discipline]

• “Informatics is the science of processing data for storage and retrieval; information science as a field.”

Artificial intelligence [The Research domain]

• "The theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

”Big Data” [The Input]

• "Big Data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.” - Gartner

Machine learning (“statistical learning”) [The Methodology]

• “Machine learning explores the study and construction of algorithms that can learn from and make predictions on data – such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs.”

"The informatics lecturer at the AI conference gave a talk on what machine learning approaches to use for Big Data."
What is informatics as a discipline?

“Medical informatics is the rapidly developing scientific field that deals with resources, devices and formalized methods for optimizing the storage, retrieval and management of biomedical information for problem solving and decision making”

Edward Shortliffe, MD, PhD

1995
...informatics is the science of information, where information is defined as data + meaning."
Remembering Ira Kalet, 1944-2015

Retired CSE adjunct professor Ira Kalet passed away last night after a long battle with cancer.

Ira joined the University of Washington in 1978 in the then newly formed Department of Radiation Oncology. Subsequently he held adjunct appointments in Computer Science & Engineering, Bioengineering, and Biological Structure, and a joint appointment in Medical Education (now the Department of Biomedical Informatics and Medical Education).
Clinical Informatics Becomes a Board-certified Medical Subspecialty Following ABMS Vote

Thursday, September 22, 2011

*AMIA to offer prep courses for clinicians who sit for Board Exam*

Washington, DC—Today, AMIA—the association for informatics professionals—announces the success of a multi-year initiative to elevate clinical informatics to an American Board of Medical Specialties (ABMS) subspecialty certified by an examination administered by the American Board of Preventive Medicine and available to physicians who have primary specialty certification through the American Board of Medical Specialties. Joining such subspecialties as pediatric anesthesiology, medical toxicology, sports medicine, geriatrics medicine, and cardiovascular disease, clinical informatics (CI) certification will be based on a rigorous set of core competencies, heavily influenced by publications on the subject.
ABMS Informatics Sub-Specialty- Very Cool!!
Biomedical Informatics in Perspective

Basic Research

Biomolecular Imaging

Biomedical Informatics Methods, Techniques, and Theories

Consumer Health
Pharmacogenomics

Health Informatics

Bioinformatics
Imaging Informatics
Clinical Informatics
Public Health Informatics

Molecular and Cellular Processes
Tissues and Organs
Individuals (Patients)
Populations And Society

Continuum with “Fuzzy” Boundaries
Technology for Innovation in Radiation Oncology

Indrin J. Chetty, PhD, Mary K. Martel, PhD, David A. Jaffray, PhD

1. Integrating radiation oncology databases across the clinic is a critical component of clinical care (45). The creation of a Virtual Clinical Trials Group for conducting retrospective research is in consideration (46). Sharing practices and outcomes will permit high mean and high variance of treatment methods, thereby improving quality (46).

2. Tools need to be created and made available for patients and physicians to discuss treatment options. The American Society for Radiation Oncology, for example, has recommended a patient decision aid tool (47). Such an approach will drive the development of more computerized decision-making tools. It is likely that patients will be able to choose treatment options based on the quality of evidence and the expected results.

3. Expertise in the informatics domain among radiation oncology professionals needs to be developed (4). The most suitable candidates with the appropriate skill sets and multidisciplinary knowledge to succeed in this space are likely medical physicists or physicians with strong computational backgrounds. Training grants for developing programs for oncology informatics will provide these individuals with the knowledge needed to support informatics research initiatives.

4. The impact of technology on radiation oncology is significant. The use of advanced imaging and treatment planning software has improved the accuracy and efficiency of radiation therapy. The use of image-guided radiation therapy (IGRT) has reduced the incidence of serious complications, increased the local control rate, and improved patient outcomes (48). IGRT is now standard of care in many institutions.

5. The integration of artificial intelligence (AI) into radiation oncology is another area of significant potential. AI can be used to analyze medical images, predict treatment outcomes, and optimize treatment plans. However, the ethical implications of using AI in radiation oncology must be considered, including the potential for bias and the need for human oversight (49).
Assessing the Training and Research Environment for Genomics, Bioinformatics, and Immunology in Radiation Oncology

Kent W Mouw MD PhD\textsuperscript{1}, Tyler F Beck PhD\textsuperscript{2}, Judith C Keen PhD\textsuperscript{2}, Adam P Dicker MD PhD, FASTRO\textsuperscript{3}

Is training in each of the three following areas sufficient for those in radiation oncology?

![Bar chart showing the number of respondents for training sufficiency in Bioinformatics, Genomics, and Immunology across different sufficiency levels.](chart.png)
Do you think that training in the following areas would allow you to advance your research career?

Assessing the Training and Research Environment for Genomics, Bioinformatics, and Immunology in Radiation Oncology

Kent W Mouw MD PhD, Tyler F Beck PhD, Judith C Keen PhD, Adam P Dicker MD PhD, FASTRO
Would you be interested in taking formal training courses in the following subject areas?

<table>
<thead>
<tr>
<th>Bioinformatics</th>
<th>Definitely not</th>
<th>Probably not</th>
<th>May or may not be interested</th>
<th>Probably</th>
<th>Definitely</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currently in residency/fellowship</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>0-7 years post-residency/fellowship</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>8+ years post-residency/fellowship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Example: Many #RadOnc residents have limited "Big Data" exposure/background

- We informally polled trainees during the MD Anderson Annual RadBio/Physics course during their informatics lecture
  - In 2018, 0/21 participants had heard of "FAIR Guiding Principles"
  - 4/21 could correctly map concepts of a "standard", "ontology", or "syntax"
  - 4/21 had heard of SNO-MED
  - 3/21 could correctly identify HL7 as a standards framework
  - 20/21 could identify DICOM, but most considered a filetype, not as an informatics standard
Box 2 | The FAIR Guiding Principles

To be Findable:
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data

To be Reusable:
R1. meta(data) are richly described with a plurality of accurate and relevant attributes
R1.1. (meta)data are released with a clear and accessible data usage license
R1.2. (meta)data are associated with detailed provenance
R1.3. (meta)data meet domain-relevant community standards
Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.⁹

3. FAIR FOR MACHINES AS WELL A PEOPLE

In eScience, two clearly separated substrates for knowledge discovery can be distinguished.

1. The actual data, which is as a rule beyond human intellectual capacity to analyse and
2. The 'Explicitome' (everything we already made explicit in text, databases and any other format to date).

- Data should be Findable
- Data should be Accessible
- Data should be Interoperable
- Data should be Re-usable.
Data Descriptor: Dynamic contrast-enhanced magnetic resonance imaging for head and neck cancers

Joint Head and Neck Radiotherapy-MRI Development Cooperative

<table>
<thead>
<tr>
<th>Data record</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical meta-data for DCE-MRI</td>
<td>Clinical meta-data for the 15 oropharyngeal cancer patients, representing our cohort. Patients, disease and treatment identifiers are detailed in this *.csv file.</td>
</tr>
<tr>
<td>data set.csv</td>
<td></td>
</tr>
<tr>
<td>ReadMe_Clinical meta-data.csv</td>
<td>Supplemental information about the headings of the columns in the clinical data file.</td>
</tr>
<tr>
<td>Arterial Input Function.csv</td>
<td>Arterial input function (AIF) extracted from DCE-MRI and required for pharmacokinetic modelling of tumors are provided.</td>
</tr>
</tbody>
</table>

Table 4. Description of data records uploaded to the figshare repository of the HPV and local recurrence prediction challenges.
NCI ANNOUNCES ONCOLOGY DATA SCIENCE FELLOWSHIP

3/13/2017

In response to the need to train oncology professionals capable of harnessing the power of big data and emerging health information technologies, the National Cancer Institute (NCI) has developed an innovative oncology fellowship program for qualified radiation oncologists.

The one-year fellowship will provide the fellow with the opportunity to develop and apply cutting-edge oncology data science knowledge to mentored research opportunities at both NCI and the Food and Drug Administration (FDA) Information Exchange and Data Transformation (INFORMED) program. Fellows will also have the opportunity to deliver clinical care at the NCI.

**Mentors:**

Kevin Camphausen, MD (NCI) [Fellowship Director]

Sean Khozin, MD, MPH (FDA INFORMED)

Anand Shah, MD, MPH (FDA INFORMED)

**Who should apply:**

- Radiation oncologists looking to develop in-depth expertise in data science, health technology, entrepreneurship and big data analytics
Fellow and Resident Radiation Oncology iNTensive Training in Imaging and Informatics to Empower Research Careers (FRONTI2ER)

- 2 Residents/Fellows admitted annually NIBIB.
- Trainees receive individualized training program.
- Trainees receive Applicants up to $10,000 for research-related expenses,
- All Trainees must commit to submit a K-level application.

Prajnan Das, M.D., M.S., M.P.H. Professor, & Residency Director co-PI

Dave Fuller, MD, PhD Ass. Professor, co-PI
Awarded FY2019...so far, in Year 1:

- 2 trainees awarded cycle 1 (both are Clinical Fellows).
- One trainee has imaging focus, the other informatics.
- Resident applicants will begin matriculating year 2.
- Expansion of the program with potential for increased non-NIH-funded trainees pending.
# Longitudinal Engagement of Pathology Residents

## A Proposed Approach for Informatics Training

**Luigi K. F. Rao, MD, and John R. Gilbertson, MD**

<table>
<thead>
<tr>
<th><strong>Table 1</strong></th>
<th>Longitudinal Engagement: FIRE Model for Resident Informatics Training</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Foundation</strong></td>
<td><strong>Immersion</strong></td>
</tr>
<tr>
<td>Didactics and online-based learning</td>
<td>Administrative group involvement</td>
</tr>
<tr>
<td>Skills and project alignment</td>
<td>Exposure to interaction with vendors and professional societies</td>
</tr>
<tr>
<td>Bioimage informatics example: lectures covering fundamentals of histology, optics, microscopy, digital imaging, and image analysis</td>
<td>Data quality example: hospital clinical decision support and computerized provider order entry group activities</td>
</tr>
<tr>
<td>Bioimage informatics example: Digital Imaging and Communications in Medicine Working Group membership</td>
<td></td>
</tr>
</tbody>
</table>

FIRE, foundation, immersion, refinement, expertise.

_AJCP_
# Career Enrichment Opportunities at the Scientific Frontier in Radiation Oncology

Reid F. Thompson, MD, PhD\textsuperscript{1,2}; Clifton D. Fuller, MD, PhD\textsuperscript{1,2}; Abigail T. Berman, MD\textsuperscript{4}; Sanjay Aneja, MD\textsuperscript{5}; and Charles R. Thomas Jr, MD\textsuperscript{2}

## TABLE 1. FiRE (foundation, immersion, refinement, and expertise through experiential learning) Longitudinal Engagement Concept

<table>
<thead>
<tr>
<th>Foundation</th>
<th>Immersion</th>
<th>Refinement</th>
<th>Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomics/bioinformatics example</td>
<td>Core ACGME curriculum; departmental didactic offerings; and scalable online curricula such as MOOCs</td>
<td>Observational shadowing exposure to laboratory or bioinformatics at home facility</td>
<td>Hands-on experiences with core facilities and researchers at home institute</td>
</tr>
<tr>
<td>Clinical informatics example</td>
<td>Core ACGME curriculum; departmental didactic offerings; and scalable online curricula such as MOOCs</td>
<td>Engagement with departmental or institutional electronic medical record, picture archiving and communication system, information technology, or other informatics specialists</td>
<td>Development of a quality assessment/quality improvement project with informatics components</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Medical physics or imaging analysis projects that intersect with informatics processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Electronic health record–based health services research</td>
</tr>
</tbody>
</table>
What about post-Rad Onc Training?

Fellowship in Clinical Informatics: Radiation Oncology Track

Clinical informatics is the subspecialty of all medical specialties that transforms health care by analyzing, designing, implementing, and evaluating information and communication systems to improve patient care, enhance access to care, advance individual and population health outcomes, and strengthen the clinician-patient relationship.

Charles R. Thomas, Jr., MD, Chair,
Unmet educational needs

• Limited time for devoted machine learning/informatics training and lack of distributed expertise will necessitate online resources (MOOCs, modular courses) as a ROECSG-supported effort
• Potential integration into current rad onc statistics education my be aided by modular exercises Greater support for specific "computational physician-scientist phenotype" by NCI/NLM/NIBIB through targeted training awards
• Development of a core leaership cadre of "first wave" junior faculty physicians through avenues such as U. Michigan Practical Big Data Workshop, ACR Data Science Institute, NCI initiatives focused on physician-scientists.
Moving forward

• How do we develop the pipeline of computational scientists who are also radiation oncologists?
  • Traditional MSTP programs train basic/translational "wet lab" scientists
  • Most Big Data expertise is ad hoc
• Need earlier integration into training for maximum efficacy (even if at a limited number of training programs)
Revitalizing the Radiology Research Enterprise

The Radiological Society of North America has embarked on a program to assist academic radiology departments in becoming more successful in their research mission. As part of this program, the Society is offering a course for chairpersons and other members of departments involved in stimulating and managing research efforts. Acquiring the understanding and the skills necessary to develop, administer, and nurture research programs of various maturity will be the focus of the course.
Can we develop Holman-pathway compatible AI-focused educational efforts leading to MS/PhD as part of residency for select trainees?

Could this be accomplished as a multi-site program?

Can this be done through existing structures (e.g. Moffitt PSOC, CTSA programs, NLM training programs) or is NCI support vital?
Next steps

• Development of a Rad Onc–focused "core curriculum" in informatics and machine learning (John Kang/Erin Gillespie) with possible ROECSG support.
• Develop curriculum into a MOOC/modular course
• Encourage AAPM TG participation by junior faculty recruitment to Big Data subcommittee (Chuck Mayo)
• Identify additional stakeholders (ASTRO, NCI, AAPM, RSNA) to define coherent vision/pathway for training needs across programs.

• IDEAS/SUGGESTIONS APPRECIATED!!
The future for enhanced performance looks good for advances in informatics, AI, and Big Data in #RadOnc, if we can build the necessary educational tools!

Please email/visit soon!

cdfuller@mdanderson.org

Caroline Chung, MD
Rad Onc MR Program Lead.