11-1-2023

Operational Ontology for Oncology (o3): A professional society-based, multistakeholder, consensus-driven informatics standard supporting clinical and research use of real-world data from patients treated for cancer

Charles S Mayo

Walter Bosch

et al.

INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

CLINICAL INVESTIGATION

Operational Ontology for Oncology (O3): A Professional Society-Based, Multistakeholder, Consensus-Driven Informatics Standard Supporting Clinical and Research Use of Real-World Data From Patients Treated for Cancer

Charles S. Mayo, PhD,[*] Mary U. Feng, MD, PhD,[†] Kristy K. Brock, PhD,[‡] Randi Kudner, MFA,[§] Peter Balter, PhD,[‡] Jeffrey C. Buchsbaum, MD,[‖] Amanda Caissie, MD, PhD,[¶] Elizabeth Covington, PhD,[*] Emily C. Daugherty, MD,[#] Andre L. Dekker, PhD,[**] Clifton D. Fuller, MD, PhD,[‡] Anneka L. Hallstrom, BS,[††] David S. Hong,[‡‡] Julian C. Hong, MD,[†] Sophia C. Kamran, MD,[§§] Eva Katsoulakis, MD,[‖‖] John Kildea, PhD,[¶¶] Andra V. Krauze, MD,[##] Jon J. Kruse, PhD,[***] Tod McNutt, PhD,[†††] Michelle Mierzwa, MD,[*] Amy Moreno, MD,[‡] Jatinder R. Palta, PhD,[‡‡‡] Richard Popple, PhD,[§§§] Thomas G. Purdie, PhD,[‖‖‖] Susan Richardson, PhD,[¶¶¶] Gregory C. Sharp, PhD,[§§] Shiraishi Satomi, PhD,[***] Lawrence R. Tarbox, PhD,[###] Aradhana M. Venkatesan, MD,[‡] Alon Witztum, PhD,[†] Kelly E. Woods, PhD,[‡‡] Yuan Yao, PhD,[*] Keyvan Farahani, PhD,[‖] Sanjay Aneja, MD,[****] Peter E. Gabriel, MD,[††††] Lubomire Hadjiiski, PhD,[*] Dan Ruan, PhD,[‡‡‡‡] Jeffrey H. Siewerdsen,[†††] Steven Bratt, BS,[§§§§] Michelle Casagni, BS,[§§§§] Su Chen, MD,[§§§§] John C. Christodouleas, MD, PhD,[††††] Anthony DiDonato, BS,[§§§§] James Hayman, MD,[*] Rishhab Kapoor, PhD,[‡‡‡] Saul Kravitz, BS,[§§§§] Sharon Sebastian, BS,[§§§§] Martin Von Siebenthal, BS,[‖‖‖‖] Walter Bosch, PhD,[¶¶¶¶] Coen Hurkmans, PhD,[####] Sue S. Yom, MD, PhD,[†] and Ying Xiao, PhD[††††]

[*]*University of Michigan;* [†]*University of California, San Francisco;* [‡]*MD Anderson Cancer Center;* [§]*American Society of Radiation Oncology;* [‖]*National Cancer Institute;* [¶]*Dalhousie University;* [#]*University of Cincinnati Medical Center;* [**]*Maastro Clinic;* [††]*Wellesley College;* [‡‡]*University of Southern California;* [§§]*Massachusetts General Hospital;* [‖‖]*James A. Haley VA Medical Center;* [¶¶]*McGill University;* [##]*National Institutes of Health;* [***]*Mayo Clinic;* [†††]*Johns Hopkins University;* [‡‡‡]*Virginia Commonwealth University;* [§§§]*University of Alabama at Birmingham;* [‖‖‖]*Princess Margaret Cancer Center;* [¶¶¶]*Swedish Cancer Institute;* [###]*University of Arkansas;*

Corresponding author: Charles S. Mayo, PhD; E-mail: cmayo@med.umich.edu

DOI of original article: http://dx.doi.org/10.1016/j.ijrobp.2023.02.058.

Disclosures: This article represents the opinion of the authors. It does not represent the opinion or policy of the National Institutes of Health of the US Federal Government or of The MITRE Corporation.

Research data are not available at this time. The O3 framework will be released into the public domain.

Acknowledgments—This work is only possible through the combined effort of a large number of people working on differing components of O3 and the multistakeholder methodology. The author list cannot adequately reflect the contributions of each of these individuals. Several authors also spent countless hours outside of the American Association of Physicists in Medicine (AAPM) Big Data Subcommittee (BDSC) meetings coordinating efforts with AAPM, American Society of Radiation Oncology (ASTRO), Canadian Association of Radiation Oncology (CARO), Common Oncology Data Elements eXtensions (CodeX), and other stakeholder groups. We do wish to highlight the efforts of a few individuals who, in addition to the BDSC meetings, also invested substantial time on additional subprojects or who may not be listed as authors. We thank David Hong, Kaley Woods, Huaizhi Geng, Sang Ho Lee, Du Wang, and Ying Xiao for the herculean effort of looking up mapping codes from O3 to SNOMED CT and National Cancer Institute (NCI) Thesaurus. George Chang, Luybov Remennick, and Lori Whiteman of the National Institutes of Health/NCI carefully reviewed assignment of O3 codes mapped to codes in NCI Thesaurus and pointed out recommended changes. During our professional society and public review cycle in July 2021, many individuals made valuable comments and corrections, which were incorporated into this version of O3. In particular, we wish to thank Laura Cervino, Drew Hope, Colleen Lawton, Chris Guy, Marie-Chantal Gagné, Ellen Kim, Jean Moran, Alberto Traverso, Emiliano Spezi, Joanna Kazmierska, Bill Nailon, Michaël Claessens, Lakshmi Santanam, Pengpeng Zhang, and Mark Phillips for their insights and suggestions. We thank Anne Hubbard for her direction on ASTRO's work with the HEART score and applications of social determinants of health measures.

This paper's review process was managed by the journal's deputy editor Salma Jabbour.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2023.05.033.

Int J Radiation Oncol Biol Phys, Vol. 117, No. 3, pp. 533–550, 2023

https://doi.org/10.1016/j.ijrobp.2023.05.033

[*****]Yale University; [††††]University of Pennsylvania; [‡‡‡‡]University of California, Los Angeles; [§§§§]MITRE Corporation; [‖‖‖‖]Varian Medical Systems; [¶¶¶¶]Washington University in St. Louis; and [####]Catharina Hospital Eindhoven

Received Jul 7, 2022; Accepted for publication May 19, 2023

Purpose: The ongoing lack of data standardization severely undermines the potential for automated learning from the vast amount of information routinely archived in electronic health records (EHRs), radiation oncology information systems, treatment planning systems, and other cancer care and outcomes databases. We sought to create a standardized ontology for clinical data, social determinants of health, and other radiation oncology concepts and interrelationships.

Methods and Materials: The American Association of Physicists in Medicine's Big Data Science Committee was initiated in July 2019 to explore common ground from the stakeholders' collective experience of issues that typically compromise the formation of large inter- and intra-institutional databases from EHRs. The Big Data Science Committee adopted an iterative, cyclical approach to engaging stakeholders beyond its membership to optimize the integration of diverse perspectives from the community.

Results: We developed the Operational Ontology for Oncology (O3), which identified 42 key elements, 359 attributes, 144 value sets, and 155 relationships ranked in relative importance of clinical significance, likelihood of availability in EHRs, and the ability to modify routine clinical processes to permit aggregation. Recommendations are provided for best use and development of the O3 to 4 constituencies: device manufacturers, centers of clinical care, researchers, and professional societies.

Conclusions: O3 is designed to extend and interoperate with existing global infrastructure and data science standards. The implementation of these recommendations will lower the barriers for aggregation of information that could be used to create large, representative, findable, accessible, interoperable, and reusable data sets to support the scientific objectives of grant programs. The construction of comprehensive "real-world" data sets and application of advanced analytical techniques, including artificial intelligence, holds the potential to revolutionize patient management and improve outcomes by leveraging increased access to information derived from larger, more representative data sets.

Introduction

The construction of large, comprehensive, "real-world" data sets and application of advanced analytical techniques, including artificial intelligence (AI), holds the potential to revolutionize patient management and improve outcomes by leveraging increased access to information derived from larger, more representative data sets. Doing so will refine the ability of providers and researchers to identify prognostic and predictive factors, offer safe treatments, and improve relevant outcomes. Comprehensive data sets, aggregated at scale through structured integration into routine practice, could facilitate novel and meaningful observations, including meeting critical needs such as characterizing covariation of control and toxicity outcomes of anticancer treatments for geographic or social determinants of health (SDOH), increasing knowledge on rare cancers and permitting benchmarking of practice performance and technical radiation therapy (RT) plan quality.

However, the ongoing lack of data standardization severely undermines the potential for automated learning from the vast amount of information routinely archived in electronic health records (EHRs), radiation oncology information systems (ROIS), treatment planning systems (TPSs), and other cancer care and outcomes databases. Standardization is necessary for the scaled assemblage of AI/analysis-ready data sets. Standardization will facilitate automating the accurate and specific extraction of data parameters from EHRs, supporting interoperable data exchange among multiple institutions and reducing "noise" in large-scale data sets. Only by incorporating data standardization into routine clinical practice will the paradigm shift from "data mining" to "data farming," with associated reduction in work needed to construct, populate, and maintain reliable databases that support clinical quality improvement and research activities.[1]

Specific domain expertise is crucial to implement large-scale data farming within a specialty, especially one as complex and technologically driven as radiation oncology. Clinicians and operators within a given specialty have the most precise understanding of which parameters are accepted as meaningful, how processes can be made feasible, and what information would be most helpful in support of quality and research efforts. Standards created by small or external groups lacking diversely represented domain expertise may omit factors crucial to clinical implementation within a specialty.[2] Development of methods validated through collective experience are material to the creation of data standards amenable to clinical and investigative use.

Prior work has demonstrated the value of an approach to creating consensus-driven information standards derived from the experience of a wide range of clinical users, professional societies, and relevant stakeholder groups. Several RT professional societies played a critical role in the development of the American Association of Physicists in Medicine's (AAPM's) task group on Standardizing Nomenclatures in Radiation Oncology (TG-263).[3] Since its release in 2017, TG-263 standards have been adopted into clinical trials and

vendor TPSs, supporting interoperable data exchange. The Global Harmonization Group recommended usage of TG-263 as part of its guidelines for organ at risk delineation for RT clinical trials.[4] Recognizing the value of RT nomenclature standardization in routine clinical practice, other professional societies, such as the Canadian Partnership for Quality Radiotherapy (CPQR), have promoted the adoption of TG-263. In 2021, CPQR released the "Guidance on the Use of Common Nomenclature in Canadian Radiation Treatment Programs," highlighting TG-263 as the foundation of future national RT "Big Data" efforts. CPQR aims to harmonize radiation oncology practice in Canada while facilitating large-scale data analysis of patient-, diagnosis-, and treatment-related factors affecting radiation oncology patient outcomes.[5,6]

We report here on a stakeholder engagement method to transparently tap the combined expertise rooted in the activities and efforts of professional societies and knowledgeable parties to specify a standardized Operational Ontology for Oncology (O3), facilitating interoperable data exchange and supporting future development of large-scale AI and analysis-ready data sets (Fig. 1). The O3 development represents the single most collaborative effort in this domain to date. It has been enabled by the collaboration of the AAPM's Big Data Subcommittee (BDSC), with representatives from the American Society of Radiation Oncology (ASTRO), the Canadian Organization of Medical Physicists (COMP), the Canadian Association of Radiation Oncology (CARO), the European Society of Therapeutic Radiation Oncology (ESTRO), and clinical trial experts from NRG Oncology. Drawing from its multiprofessional, multinational, and multi-institutional members' clinical, research, and informatics expertise and using a consensus-driven approach, the AAPM's BDSC identified key clinical elements, attributes, value sets, and relationships deemed necessary for standardization to support RT clinical practice and research applications. In recognition of its genesis with the professional societies cited, the ontology was originally named the "Operational Ontology for Radiation Oncology," even though the comprehensive ontology extends to all parts of cancer care. To reduce confusion in explaining the scope of the ontology, the "R" was dropped, leaving the "Operational Ontology for Oncology" (O3).

A nomenclature identifies a set of standardized names to improve consistency; for example, AAPM TG-263 identifies standardized names for structures segmented as part of treatment planning. A taxonomy identifies a hierarchy of categorizations, for example, the Linnaeus system for subdividing categorizations of plant types.

An ontology is a philosophical representation of a body of knowledge that identifies the concepts and interrelationships. It provides the basis for a coordinated reasoning about that body of knowledge.[7] A challenge for clinicians applying the term "ontology" is that it tends to be abstract, and it is difficult to provide practical examples. In this work, we use the term "operational ontology" to clarify that all concepts, value sets, relationships, and so on are defined for practical application to facilitate use and improve consistency of clinical and research operations as well as linkages to other information coding systems. The operational ontology combines concepts of nomenclature and taxonomy with ontology to develop operational standardizations. The operational ontology is presented in a form that end-users can comprehend and contextualize in their own practice.

Note that this use of "operational ontology" is distinct from a "sematic-web ontology" that expresses concepts and relationships in a computer readable format that is not organized for end-user development and consumption. For example, sematic-web ontologies generally do not provide standardized value sets, linkages to clinically used categorizations, coding systems, definitions, type definitions, and so on. Formal sematic-web ontologies (ie, "formal ontologies"), such as the Open Biological and Biomedical Ontologies (OBO) Foundry, try to cover knowledge domains using formal definitions while being interoperable with coordinated use of terms.[8] Examples of semantic-web ontologies can be found at the Bioportal or OBO Foundry websites.[9] In practice, semantic-web ontologies leave operational gaps for clinical practice, are not developed in the context of multiple professional society efforts with peer review and public comment, and are not developed with the express aim to make real world data interoperable. These gaps can be filled with a professional society-based, operational ontology to improve consistency in creating interoperable, AI-ready data sets needed to support clinical and research efforts. Development of an operational ontology for cancer care provides a strong basis for subsequent development of formal semantic-web ontologies.

The stakeholder engagement method was developed alongside the creation of the O3. It aimed to support incremental engagement with a wide range of stakeholders, allowing for downstream optimization by addressing several clinical and research objectives shared within professional societies to expand the codified sets of disease site-specific elements. Multiple additional stakeholders were engaged, including government agencies and vendors, underscoring the value of a collaborative approach to developing comprehensive and standardized ontologies. The O3 has been endorsed by AAPM, ASTRO, CARO, COMP, and ESTRO.

The initial version of O3, described herein, includes many disease-site independent concepts common for all patients with cancer and a smaller set specific for prostate cancer. These include imaging, pathology, medical oncology, surgery, and radiation treatment. The O3 methodology is currently being applied/adapted to disease site-specific concepts for head and neck cancers and breast cancer through the AAPM BDSC, with lung and remaining sites to follow. The clinically oriented O3 lays the foundation for informatician-oriented formalized ontology tools following the principles established by the OBO Foundry.[8] It supports multi-institutional federated and centralized databases and registries by including a standardized, O3-driven database schema and tables for enumerating standardized values. This provides a professional society-driven pathway toward interoperable, AI-ready, multi-institutional databases.

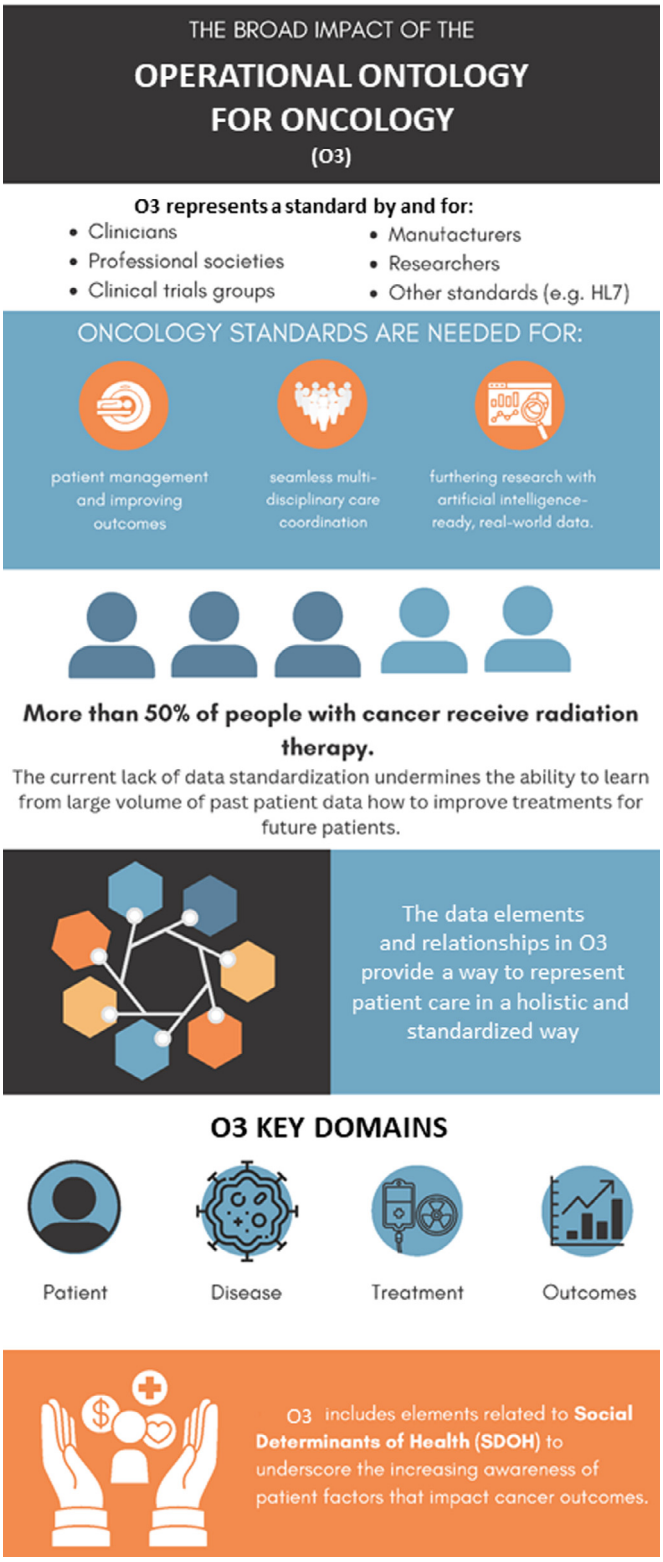


Fig. 1. A professional society-based, consensus-driven approach was taken in developing the Operational Ontology for Oncology to address standardization gaps that undermine ability to aggregate and learn from real-world data, supporting clinical practice improvement and research.

Methods and Materials

The AAPM BDSC is currently composed of a diverse international group of stakeholders, including 38 members: physicists (n = 20), physicians (n = 13), informaticians (n = 3), and other representatives (n = 2). The BDSC includes members of several professional societies: AAPM (n = 24), ASTRO (n = 26), CARO (n = 1), COMP (n = 2), ESTRO (n = 6), American Society of Clinical Oncology (ASCO) (n = 2), Radiological Society of North America (RSNA) (n = 8), and American Medical Informatics Association (AMIA) (n = 3), as well as members of NRG Oncology (n = 8), the National Institutes of Health (NIH) (n = 4), National Cancer Institute (NCI) (n = 3), IROC (n = 1), Integrating the Health Care Enterprise Radiation Oncology (IHE-RO) (n = 2), United States Veterans Administration (VA) (n = 2), Southwestern Oncology Group (n = 3), and Digital Imaging and Communications in Medicine (DICOM) (n = 1).[10] The number of members on the committee increased over time from the initial set of 23, as professional societies appointed representatives (eg, ASTRO, n = 3) and as additional specialty area gaps were identified with resultant need to augment the base of domain experts (eg, technologies such as heavy particle therapy, brachytherapy, diagnostic radiology, multinational practitioners).

Engagement with external initiatives

The 21st Century Cures Act promoted implementation of Health Level 7 (HL7) Fast Healthcare Interoperability Resources (HL7 FHIR) as the basis for improving patient access to their health care data and interoperable exchange of data among their providers. It requires the creation of FHIR based application programming interfaces to align industry efforts around interoperability and provide broader access to health care data.[11,12] HL7 FHIR supersedes HL7. HL7 is a set of standards widely used by hospital systems to transmit health care data. The HL7 FHIR standard overcomes significant gaps in HL7 by providing means to extend it with medical domain-specific coding. Development of medical domain-specific FHIR tags is being carried out as part of the minimal Common Oncology Data Elements (mCODE) and Common Oncology Data Elements eXtensions (CodeX) projects organized by HL7 and The MITRE Corporation. The MITRE Corporation is a not-for-profit organization that works in the public interest across federal, state, and local governments, as well as industry and academia.

CodeX is a community accelerator focused on interoperable data modeling and applications that are integrating and testing mCode implementation based on the HL7 FHIR. Within CodeX, the RT treatment data (RTTD) use case team began work on a treatment summary use case in March 2021. RTTD use case members and ROIS and EHR representatives worked with HL7 and IHE-RO to develop HL7 FHIR frameworks. BDSC members are leaders in both CodeX and RTTD use case efforts. With the strong RT focus, the elements, attributes, value sets, and relationships of O3 are a superset of the elements addressed in mCODE and CodeX, informing development of those standards. AAPM and ASTRO are founding members of CodeX, with BDSC representatives leading the RTTD CodeX use case project started in March 2021.[13] BDSC representatives in RTTD successfully lobbied for IHE-RO medical physics inclusion in RTTD use case members in addition to radiation oncologists. AAPM also sponsored COMP and the Society for Imaging Informatics in Medicine to join as CodeX and RTTD members.

Development of O3 in AAPM BDSC intersects with other efforts. BDSC communicated with members of the Office of the National Coordinator for Health Information Technology, proposing O3-driven extensions to the United States Core Data for Interoperability standards. Mappings of O3 concepts to VA-ASTRO collaborations in developing quality measures were supported with BDSC members who were part of both efforts.[14]

To promote interoperability of O3 with other coding systems, for example, SNOMED CT[15] and NCI Thesaurus, BDSC enumerated and mapped O3 attributes to attributes in these other systems. NCI Thesaurus is a framework coding system for medical concepts used in clinical trial research. Similar to semantic-web ontologies, gaps and challenges in structure prompted recognition that this effort carried out by the professional society-based stakeholders would be of general benefit. BDSC engaged with leadership of the NCI Thesaurus to validate mappings and homologate the O3. SNOMED CT is a coding system for multilingual health concepts. It has been designated as a national standard used by Office of the National Coordinator for Health Information Technology and in HL7 FHIR. BDSC created mappings to SNOMED CT codes and identified gaps. It worked in conjunction with CodeX in identifying and requesting SNOMED CT codes for missing concepts.

Intersections of O3 with DICOM were explored with DICOM leaders. Members of BDSC include leaders in IHE-RO and the AAPM DICOM Working Group. O3 was presented and discussed at meetings of the DICOM working group at the AAPM annual meeting to promote mutual understanding and development of mutually beneficial standardizations.

Development process

AAPM's BDSC was initiated in July 2019 and initially held monthly videoconference meetings to explore common ground from the stakeholders' collective experience of issues that typically compromise the formation of large inter- and intra- institutional databases from electronic records. The Radiation Oncology Translational Research Ontology detailed in the appendix of a 2018 publication by Mayo et al[16,17] from the conference proceedings of the 2017 Practical Big Data Workshop formed the starting point for the

O3. Together with the prior work of several members in creating ontologies and clinical databases, the group's experience formed a strong basis for development.[1,18-21] The group decided to focus on prostate cancer as the initial use case for identifying attributes of key elements generalizable across disease site groups (eg, RT course), as well as disease site-specific subsets (eg, diagnosis and staging components for prostate cancer). Prostate cancer was chosen because of its high population incidence and public health relevance, with well-defined control and toxicity outcomes, as well as to reinforce joint efforts of ASTRO and the VA on prostate treatment quality measures at that time. The BDSC then increased the frequency of meetings to biweekly in April 2020 to enable rapid progress in detailing the O3.

An iterative-deliberations approach was used to incrementally construct and refine the ontology. Versions were recorded in standardized format Word document with standardized formatting delineating all aspects of key elements, attributes, and relationships. The Word document was maintained on a shared web directory and new versions with changes were e-mailed to members. An application was created to parse iterations in the Word document and render it into a JavaScript Object Notation (JSON) file used to iteratively populate the O3 website. This provided a different format for inspection from the Word document to facilitate narrowing focus to specific sections of interest.

The group iteratively worked through each key element, attribute, value set, relationship, and prioritization, asking for objections and identifying gaps or omissions. Conversation on an item continued until there were no further issues or objections raised. If issues were subsequently identified, the group returned to the discussion of the item. The group worked until consensus was reached. Members not in attendance could reintroduce topics voted on previously. For highly specific, physician-centric processes, smaller subgroups were identified and carried on the same process and then brought recommendations back to the larger group.

For attributes where there was consensus among clinicians on common value sets encountered in practice, for example, imaging modalities used for M staging, these were enumerated, and the list was identified as extensible with other values. For those without common value sets, a Delphi process was employed to obtain consensus.[22,23] Survey recipients were members of ASTRO disease site specialty groups, BDSC, and other researchers, and clinicians identified as actively engaged in standards development, deployment, and use.

For final recommendations of prioritization groupings for use in recommendations for clinical focus, scores were calculated based on the prioritizations assigned for key elements and attributes. Tables were constructed grouping attributes. These were iteratively reviewed with priorities for key elements and attributes refined to accord with group consensus for areas of focus.

The BDSC adopted an iterative, cyclical approach to engaging stakeholders beyond its membership to optimize the integration of diverse perspectives from the community (Fig. 2A). The group took an approach during the project years (Fig. 2B) permitting the addition of new data elements, as clinical and technical concepts evolved over time, to avoid obsolescence. This iterative process included collating features deemed essential for supporting most clinical and research efforts in cancer care. The ontology's concepts underwent several iterations to optimize a hierarchical framework for data capture and reasoning from clinical records. An open-access website was created to enable public comment and to support wider distribution.[24] The BDSC solicited targeted, multistakeholder review and feedback beyond its membership in July 2021 and again in April 2022, with expansions and refinements made based on feedback received. Targeted engagements and feedback were also solicited from AAPM and ASTRO leadership, as well as CPQR contacts at the 46 RT centers across Canada, CARO, COMP, ESTRO, major vendors of electronic records systems, and The Mitre Coporation and HL7 leadership on CodeX projects.

Results

O3 version 1.0

The first version of O3 identified 42 key elements and 359 attributes, 144 value sets, and 155 relationships. Deliberations resulted in the ranked relative importance of key elements and attributes by merging clinical significance with the likelihood of availability in electronic records or the ability to modify routine clinical processes to permit aggregation. The subset of prostate cancer-specific attributes and value sets were a small fraction of the totals, 8 and 4, respectively. The group used rankings to subgroup the attributes into 5 priority rankings to facilitate implementation. Table 1 characterizes the first 3 priority groupings of 24, 22, and 65 attributes, respectively. The remaining priority groupings are tabulated in Appendix E1. The complete O3 is provided on the website, including links to download the complete set as a spreadsheet or in JSON format for electronic processing.[25] Figure 3 illustrates a fully detailed attribute (disease site response) as presented on the website.

Value set standardization

The group composed a standardized list of minimum sets of values that could be consistently implemented across electronic record systems and improve AI readiness by reducing the need to recategorize inputs. For each attribute identified, BDSC investigated if value sets existed that were widely recognized as an existing standard in clinical or research settings. These were listed as reference systems if the group noted more than one. The group then recommended value set items drawn from these standards (eg, American Joint Committee on Cancer for staging). Value sets for attributes were designed anticipating eventual direct use in clinical radiation oncology information system interfaces to support consistency and quality.

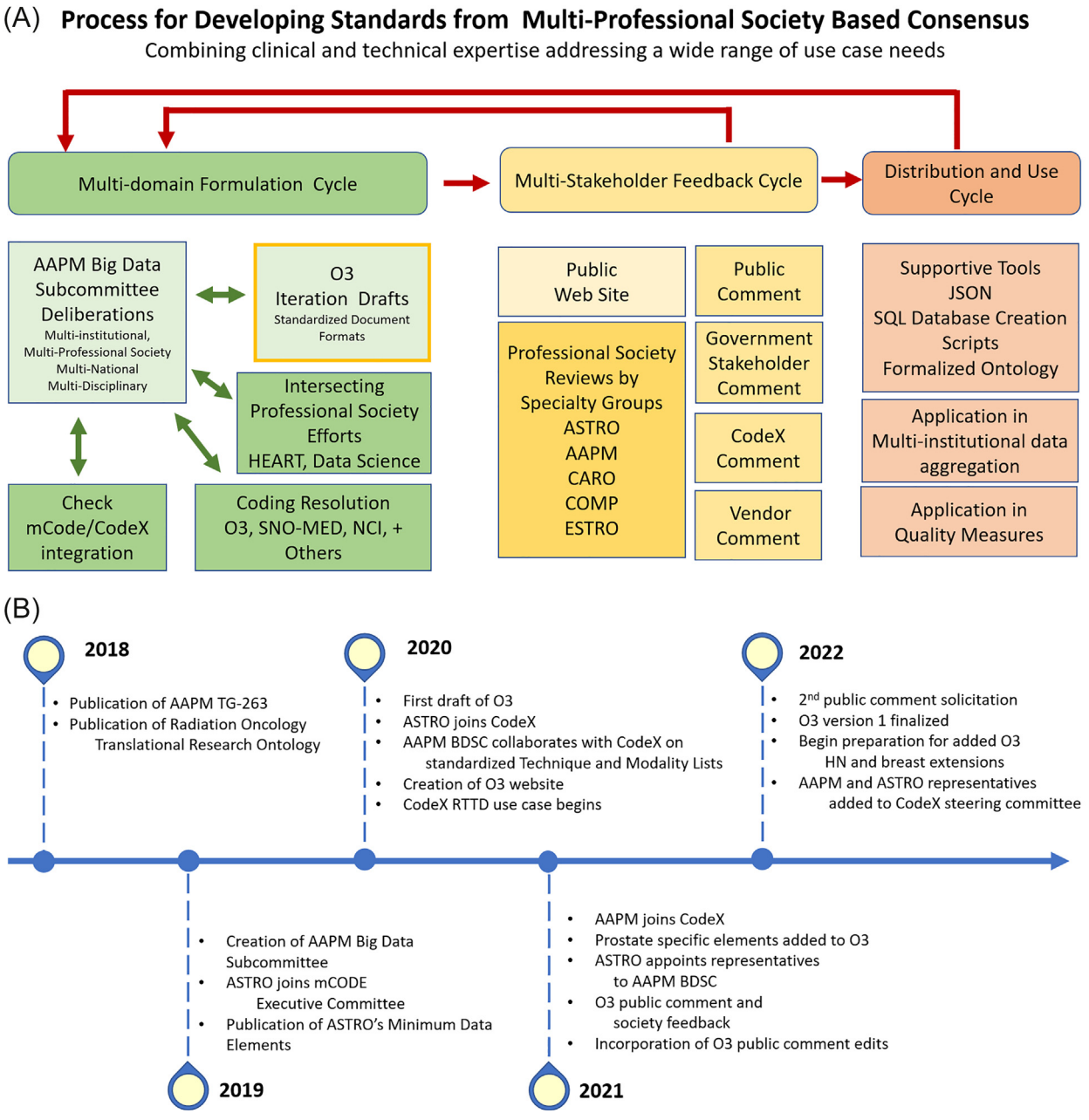


Fig. 2. (A) An iterative, progressive engagement method, which included multiple health care electronic systems vendors, government agencies, specialty groups within professional societies, institutions, and the public that were engaged with a publicly accessible website for collecting comments and direct engagement with representatives to ensure incorporation of multi-stakeholder perspectives. (B) Building on prior work from 2018, the work was carried over multiple years with progressive engagement of stakeholders.

Relationship quantification

High priority relationships between key elements that need to be captured were categorized. The existence of child key elements were predicated upon existence of the parent key element in "child element of" and "parent element of" categorizations. Elements that exist independently but for which the relationship should be captured if it exists were categorized as "is associated with."

As current radiation oncology information systems are constructed and used in practice, there is high probability of accurately capturing parent-child relationships. "Is associated with" relationships may be less reliably captured given current workflows and systems.

For example, a patient key element (parent) will be linked to each instance of a diagnosis and staging (child) record. A patient may have many diagnosis and staging records, only some of which are associated with an RT course. Ability to

Table 1 Attributes recommended as first, second, and third priority for efforts to improve systematic aggregation for all patients

First priority for aggregation efforts					
Key element	**Attribute**				
Patient	Patient MRN	❖		★	
	Anonymized patient identifier			⊙	
	Patient identifier data set			★	
	Date of birth	❖		★	
	Year of birth			★	
	Date of death	❖		★	
	Age at date of death			★	
	Sex at birth			★	
Patient information	Vital status			★	✓
	Postal code of domicile	❖	⌘	★	✓
Diagnosis and staging	Staging system			☆	✓
	ICD version			★	✓
	ICD code			★	✓
	Histology			☆	✓
	Staging T category (eg, clinical, pathologic)			★	✓
	Staging T			★	✓
	Staging N			★	✓
	Staging M			★	✓
Diagnosis and staging: treatment overview	Had radiation therapy as part of treatment			★	✓
	Had systemic therapy as part of treatment			★	✓
	Had surgical procedure as part of treatment			★	✓
Diagnosis and staging: prostate components	Had prostatectomy			☆	✓
	Prostatectomy margin status			☆	✓
	Disease status			☆	✓
Second priority for aggregation efforts					
Key element	**Attribute**				
Patient	Cause of death			⊙	
	Cause of death attributable to treatment			⊙	
	Patient-reported race		⌘	★	
	Patient-reported ethnicity		⌘	★	
	Gender identity		⌘	⊙	✓
	Has tobacco use history			★	✓
	Height			★	✓
	Weight			★	✓
	Address of domicile	❖		★	✓
	Address of primary care physician	❖		☆	✓
	Insurance		⌘	⊙	✓
	Disability		⌘	⊙	✓
	Caregiver of disabled person status		⌘	⊙	✓
	Education level		⌘	⊙	✓
	Primary language spoken at home		⌘	⊙	✓

(*Continued*)

Table 1 (*Continued*)

First priority for aggregation efforts				
Key element	**Attribute**			
Diagnosis and staging	Pathology molecular test name		☆	✓
	Pathology molecular test result		☆	✓
Diagnosis and staging: prostate components	PSA value		★	✓
	Primary Gleason score at biopsy		★	✓
	Secondary Gleason score at biopsy		★	✓
Patient treatment outcome	Evaluation basis of disease status		☆	✓
	Site of recurrence		☆	✓
Third priority for aggregation efforts				
Key element	**Attribute**			
Provider-reported toxicity	Toxicity coding system		☆	✓
	Toxicity measure		☆	✓
	Toxicity value		☆	✓
PRO	PRO instrument		⊙	✓
	PRO question ID		⊙	✓
	PRO question		⊙	✓
	PRO question response		⊙	✓
Performance score	Scoring system		☆	✓
	Performance score value		☆	✓
Radiation therapy course	Radiation therapy course facility		★	✓
	Radiation therapy course facility postal code	⌘	★	✓
	Number of sessions in course		★	✓
	Course involves reirradiation		☆	✓
Radiation therapy course: target dose	Target volume		☆	
	Target volume dose		☆	
	Dose unit		☆	
Radiation therapy prescription	Name of radiation therapy prescription		☆	✓
	Number of treatment sessions		★	✓
	Number of sessions per day		★	✓
	Number of sessions per week		★	✓
Radiation therapy prescription: target dose	Target volume		☆	
	Target volume dose		☆	
	Dose unit		☆	
Radiation therapy prescription: dose objective	Structure type		⊙	
	Structure name		⊙	
	Dose objective		⊙	
	Objective condition		⊙	
	Value		⊙	
Treated radiation therapy plan	Plan name		★	
	Number of fractions intended		★	
	Number of fractions treated		★	

(*Continued*)

Table 1 (*Continued*)

First priority for aggregation efforts				
Key element	**Attribute**			
	Adaptation or revision		☆	
	Modality		★	
	Dose delivery category		☆	
	Is plan representing cumulative phase		☆	
	Is plan representing cumulative course		☆	
	Is plan sum		★	
Radiation therapy treated plan: target dose	Target volume		☆	
	Target volume dose		☆	
	Dose unit		☆	
Treatment plan details: external beam radiation therapy	Technique		★	
	Delivery device		★	
	Energy		★	
Treatment plan details: brachytherapy	Technique		★	
	Delivery device		★	
	Energy		★	
Treatment plan details: radiopharmaceutical	Technique		☆	
	Energy		☆	
Treatment plan details: hadrons	Area coverage method		☆	
	Robust optimization		☆	
	RBE basis		☆	
	RBE		☆	
	Used repainting		☆	
Treatment plan field details: hadrons	Field name		☆	
	Range shifter		☆	
Systemic therapy course	Systemic therapy course facility		★	✓
	Systemic therapy course facility postal code	⌘	★	✓
	Systemic therapy course type		☆	✓
Systemic therapy cycle	Systemic therapy cycle name		★	✓
	Cycle status		★	✓
Systemic therapy cycle drugs used: chemotherapy	Drug name		★	
Interventional procedure	Interventional procedure facility		★	✓
	Facility postal code	⌘	★	✓
	Interventional procedure type		☆	✓
	Is preirradiation		★	✓

Attribute values constrained by PHI (❖) and those supporting measurement of SDOH (⌘) are indicated. Values judged implementable with electronic (★) or manual (☆) extraction or likely missing (⦿) from current clinical record systems are indicated. Values for longitudinal elements (✓) record date and time as well as high precision patient age. Fourth and fifth priority tables are provided in Appendix E1.

Abbreviations: ICD = International Classification of Diseases; ID = identification; M = metastasis; MRN = medical record number; N = node; PHI = patient health information; PRO = patient-reported outcome; PSA = prostate-specific antigen; RBE = relative biologic effectiveness; SDOH = social determinants of health; T = tumor.

quantify the relationships automatically electronically between diagnosis and staging and RT courses is dependent on the current capabilities of the EHR, ROIS, and TPS systems in use and how they are implemented in clinical practice.

The Delphi process was required for a small number of high priority attributes for critical concepts: disease response categorizations, relative prioritization of provider-reported outcomes, and relative prioritization of toxicities. Members of

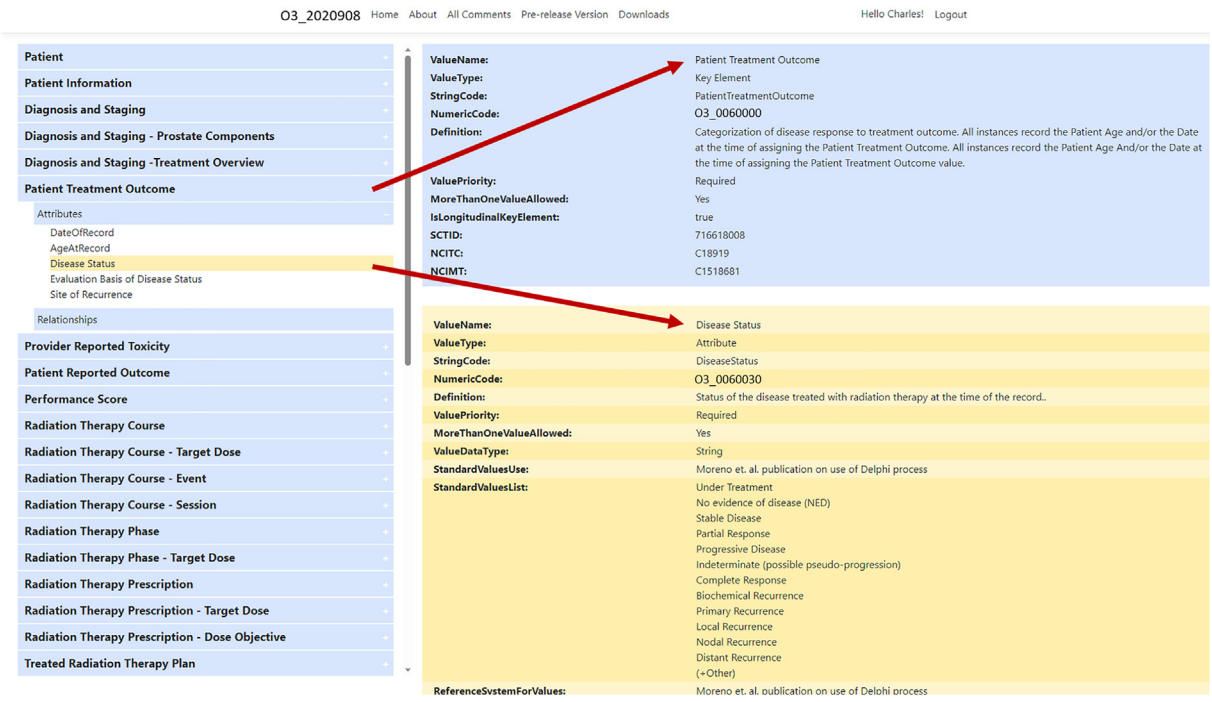


Fig. 3. An Operational Ontology for Oncology (O3) website screenshot illustrating the details of 1 of 5 attributes under the key element of patient treatment outcome. The O3 is formatted in a user-friendly drop-down menu with each key element listed to the left of the page (eg, patient treatment outcome), providing users the ability to access additional information through menu expansion, including value name (eg, disease status), value type (attribute), definition, and list of standard values with reference for such standards along with O3 codes for the value in comparison to current identifiers in SNOMED CT (SCTID), NCI Thesarus (NCITC), and NCI Metathesarus (NCIMT). Relationships that should also be tracked are tabulated.

BDSC and of ASTRO disease site specialty groups were engaged in the Delphi process using iterative survey methods and discussion to identify consensus-based recommendations.

Static versus longitudinal key elements

Key elements were classified as static or longitudinal. Longitudinal key elements have attributes that may change over time. For example, sex at birth is a static attribute, but gender identity is a longitudinal attribute. Longitudinal key elements, such as date and time, patient age as a decimal value, or both, are captured with attribute values. Recording patient age at the time of an event, relative to midnight on the date of birth, with resolution of 3, 4, and 6 decimal places allows resolving time differences of days, hours, and minutes respectively, while also meeting patient health information requirements for patients younger than 89.[16] O3 can support research scenarios with a finer resolution of time differences below the millisecond range. Our approach readily scales to this use case by increasing the number of decimal places used without requiring the recording of additional fields.

Concept definition and consistency

Lack of consistency in specific meanings for some key concepts used in describing RT was a significant barrier in standardizing use of several oncologic concepts, especially pertaining to the number of treatment courses, dose and fractionation, and outcomes. This lack of consistency affects several aspects of our field and undermines the ability to automate electronic aggregation to create and combine curated data sets. The BDSC iterated through a wide range of clinical use cases, working closely with the CodeX RTTD to identify consistent definitions for RT course, phase, sessions, and plan fractions, including multimodality treatments and plan adaptation.

For example, the concept of "phase" had lacked an "operational" standardized meaning implementable within radiation oncology information systems. Working in the context of multivendor participation, the BDSC and RTTD teams removed ambiguity. Phase was defined specifically as a concept for grouping treated plans using the same technique and modality to treat the same set of target volume dose-per-fraction pairings. Similarly, course was defined to group all treated plans from first delivery of radiation in the course until the physician believes they have treated all body sites in need of RT. A session within a course was identified as the period between when the patient enters the treatment room until they leave it and was used to group all plan fractions treated during the session. Use of decimal values for sessions allowed grouping when a session was interrupted (eg, 3.0 and 3.1 for initial and resumed treatment after a midtreatment multileaf collimator motor failure and replacement).

Figure 4 illustrates the framework of definitions for (1) prostate, (2) bilateral breast with adaptation, and (3) oligometastatic sites with asynchronous starts and response timing gaps. These definitions were adopted as standards and incorporated into HL7 FHIR version 2.0.0, with the cooperation of EHR, ROIS, and TPS vendors, and have recently (May 2022) been included in IHE-RO interoperability testing of vendor systems.

Historically, some standards systems have evolved as a patchwork of specific individual use cases rather than a comprehensive view of all information needed within a domain. Cross mapping-enumerated O3 attributes highlighted gaps in several areas for SNOMED CT concepts and in the NCI Thesaurus. O3 established a hierarchy of key elements, attributes, and value sets whereas BDSC engaged with NCI to share findings and coordinate mapping efforts. O3 contained 16 key elements and 54 attributes that did not have corresponding matches in SNOMED CT. For NCI, there were 10 unmatched key elements and 54 unmatched attributes.

Database schema

To facilitate adoption into research applications, the operational ontology was rendered as a relational database. Scripts were created to automate creation of an instance of the Structured Query Language (SQL) O3 database. Key elements were encoded as tables and attributes as fields within the SQL O3 database. SQL key fields were created to quantify relationships identified in O3. To simplify adoption, only parent-child relationships were required, allowing flexibility in encoding "is associated with" relationships. This flexibility is important when there may be missing data from specific data sets. The key elements, attributes, value sets, and coding are stored in a table of the SQL O3 database. These may be subsequently used programmatically in user interfaces to create drop-down lists to facilitate use of standard values and encoding. This facilitates later refinement of standard value sets. The approach is intended to allow flexibility for aggregating data into relational databases without adding rigidity by over specifying constraints beyond the primary parent and child relationships.

Intersections with other stakeholder groups: Treatment summaries

The RTTD use case focused on treatment summaries from a small set of items identified in previous ASTRO white papers.[19,20] Informed by the superset of elements in O3, the RTTD and BDSC groups identified a standardized operational definition for the treatment phase and enumerated values for technique and modality. The CodeX RTTD use case team successfully lobbied for resolving discrepancies and extending values in publicly available SNOMED CT codes.[26] The ASTRO, ESTRO, and AAPM-endorsed TG-263 nomenclature formed the basis of target volumes used by the BDSC and RTTD use case groups to collaboratively identify extensions to TG-263, adding certain target volume designations to handle specific radiation oncology use cases. A canonical mapping of TG-263 nomenclature terms to SNOMED_CT codes and qualifiers was developed by RTTD and incorporated into mCODE.[27]

Intersections with other stakeholder groups: SDOH

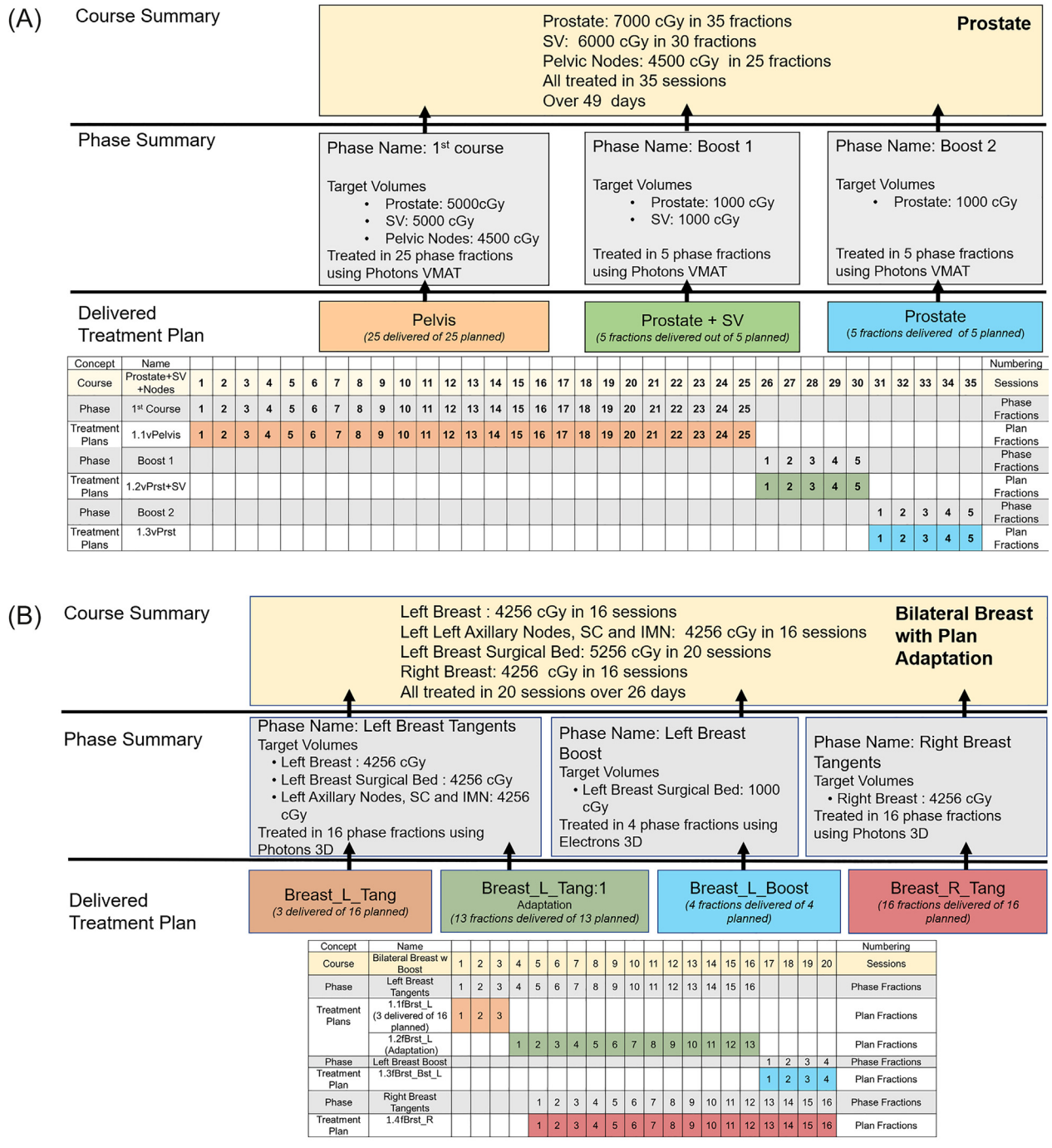
Increasing awareness of the importance of SDOH, patient ability/disability, and access to care as connected to cancer outcomes has emphasized the capture of robust data amenable to analysis that can inform policy change to improve cancer health equity.[24,28-33] In addition to traditional SDOH measures, BDSC successfully lobbied to include disability in the United States Core Data for Interoperability[34] version 2 and is lobbying for status as primary caregiver of persons with disability in version 3.

BDSC prioritized aggregating attributes quantifying social determinants of health to support modeling effects on outcomes and guide public health policy. O3 supports ASTRO's proposal of the Health Equity Achievement in Radiation Therapy (HEART) framework to capture and address health care disparities in radiation oncology.[33] ASTRO recommended that the HEART assessment be included in the proposed Radiation Oncology Alternative Payment Model to provide additional funding for wrap-around services to address health disparities, such as patient care navigators, personal support systems, and access to transportation and housing. Data associated with episodes of care with a HEART payment could be collected and used to determine the effectiveness of HEART interventions. The assessment and data would inevitably assist in determining what causes these disparities and understanding what interventions are most effective to narrow the gaps.

Developing the O3 in conjunction with the CodeX RTTD use case provided opportunities to engage with government agencies, including the American Community Survey, to identify SDOH value sets that could be implemented with minimal clinic burden. The engagement method placed a priority on clinicians identifying value sets that could reasonably be incorporated into their clinical practice. Without this engagement process, these value sets may have been determined without critical input from our professional societies.

Intersections with other stakeholder groups: DICOM

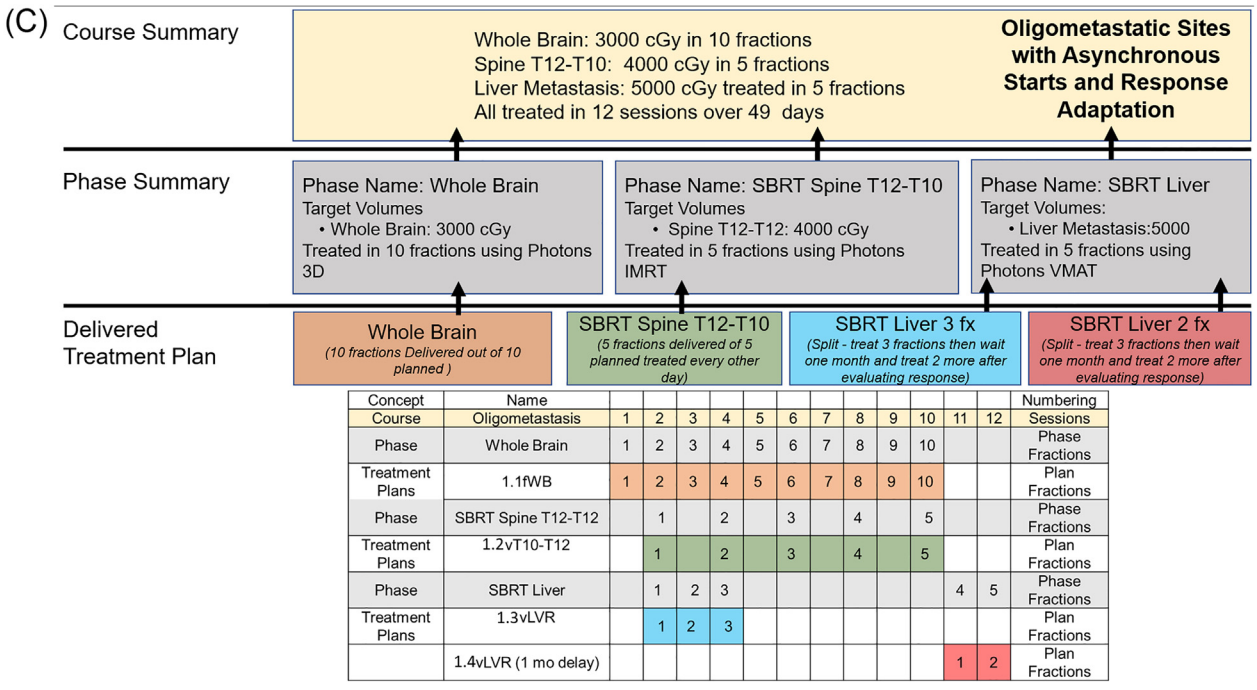
By design, the committee included members who were also primary leaders in design and implementation of DICOM, IHE-RO, and HL7 FHIR to avoid redundancy and promote dissemination of results from O3 to support those efforts as they continue to evolve. DICOM, HL7, and HL7 FHIR operate at the transport layer operationalizing transactional



Concept	Name	Numbering
Course	Prostate+SV +Nodes	Sessions
Phase	1st Course	Phase Fractions
Treatment Plans	1.1vPelvis	Plan Fractions
Phase	Boost 1	Phase Fractions
Treatment Plans	1.2vPrst+SV	Plan Fractions
Phase	Boost 2	Phase Fractions
Treatment Plans	1.3vPrst	Plan Fractions

Concept	Name	Numbering
Course	Bilateral Breast w Boost	Sessions
Phase	Left Breast Tangents	Phase Fractions
Treatment Plans	1.1fBrst_L (3 delivered of 16 planned)	Plan Fractions
	1.2fBrst_L (Adaptation)	Plan Fractions
Phase	Left Breast Boost	Phase Fractions
Treatment Plan	1.3fBrst_Bst_L	Plan Fractions
Phase	Right Breast Tangents	Phase Fractions
Treatment Plan	1.4fBrst_R	Plan Fractions

Fig. 4. Illustration of radiation therapy treatment concepts of radiation therapy course, phase, plan, session, and fraction using 3 clinical examples ordered by increasing complexity. (A) Prostate cancer, including 3 treatment plans delivered to increasingly focused targets of the pelvis, prostate/seminal vesicles, and prostate in 3 sequential phases for a total of 35 fractions/sessions. (B) Bilateral breast cancer, including 3 treatment plans delivered to left (L) breast/axilla, boost site of L breast, and right (R) breast. In this case, there is a fourth plan because of adaptation required of the L breast/axilla plan, which had only 3 fractions delivered of the 16 planned. Although the L breast/axilla and L breast boost phases were treated sequentially, the phase of R breast plan was treated concurrently (in the same sessions) with those of the contralateral treatments. (C) Oligometastatic disease, including a total of 4 treatment plans delivered over a total of 12 sessions. The 2 plans within the liver phase were treated asynchronously. After evaluation of treatment response, the second liver plan of 2 fractions was delivered 1 month after the first liver plan of 3 fractions, which had been treated concurrently with the whole brain (10 fraction) and spine (5 fraction) plans.



Concept	Name													Numbering
Course	Oligometastasis	1	2	3	4	5	6	7	8	9	10	11	12	Sessions
Phase	Whole Brain	1	2	3	4	5	6	7	8	9	10			Phase Fractions
Treatment Plans	1.1fWB	1	2	3	4	5	6	7	8	9	10			Plan Fractions
Phase	SBRT Spine T12-T12		1		2		3		4		5			Phase Fractions
Treatment Plans	1.2vT10-T12		1		2		3		4		5			Plan Fractions
Phase	SBRT Liver		1	2	3							4	5	Phase Fractions
Treatment Plans	1.3vLVR		1	2	3									Plan Fractions
	1.4vLVR (1 mo delay)											1	2	Plan Fractions

Fig. 4. Continued.

communication, that is, how is a treatment record sent from one system to another. DICOM is primarily used to transport imaging and radiation dose information. HL7 and its successor HL7 FHIR are used to transport all other health record information, for example, laboratory values. O3 operates at a higher, semantic layer identifying which specific items of information and value sets need to be stored and made transmittable, including which interrelationships between key elements must be quantified in those transmissions. When the transport layer is developed ahead of reference to a semantic layer developed by professional society consensus, then the transport protocols can miss important connections needed to fully convey concepts. For example, the detailed clinical, procedural, and conceptual connections between treatment plans, phases, courses, and doses to target volumes provided the foundation by vendors to create the HL7 FHIR connections so that treatment summary information was correctly transported between systems. Because leadership from these groups was included in the years-long development by design to promote dissemination of the work, some O3 concepts have subsequently been incorporated into those efforts.

Discussion

An ontology represents a common framework for categorizing key elements and interrelationships of concepts in a specific domain with a standardized vocabulary supporting sharing, reusing data, and enabling AI applications and computer reasoning.[7] Ontologies have a strong foothold in genomics.[35-38] Narrow scope ontologies have been gradually investigated by small groups of investigators in health care.[39-43] Challenges faced in achieving the multi-institutional, multistakeholder consensus required for comprehensive health care domain-specific ontology development have slowed their adoption and wider application.

A few nonprofessional, society-based groups are working to create standardized value sets supporting data exchange. For instance, the Observational Health Data Sciences and Informatics constructed the Observational Medical Outcomes Partnership, a broad clinical data model, which has been applied successfully to facilitate general clinical data science efforts across institutions.[44-48] However, Observational Medical Outcomes Partnership and other similar standards remain limited by their more general scope, and therefore are unable to address many narrow-scope oncology-centric problems. Lack of quantification of important relationships between key elements undermines the likelihood of design implementations in clinical systems to ensure the ability to correlate and quantify dependencies. These gaps leave a need for more granular, domain-specific data models in oncology and radiation oncology.

"Operational" is a key component of O3. O3 was developed by clinical and technical radiation oncology domain experts for implementation within the electronic systems used in routine practice that govern clinical workflows. Leveraging the team's expertise in the construction and use of metrics used in quality, safety, accreditation, billing, and research, O3 was designed to support these applications. The structure and value sets were created to support the design of consistent user interfaces within electronic systems to subsequently enable automated data extractions and reduce manual effort to implement.

By focusing on the development of a functional ontology supporting both clinical and research objectives, we have developed an engagement method for multistakeholder, consensus-driven standardized ontology development, engaging experts from key professional societies. Our goal is for this to support the creation of AI and analytics data sets from routine clinical practice, real-world data. This will facilitate a wide range of uses, including verification of existing hypothesis-driven studies, support for reproducible analysis (model, data, or both model/data driven), data aggregation and pattern discovery, and correlative causality analysis. O3-based data standardization within RT electronic systems could provide a framework for integrating quality measures with billing codes, facilitating efficacy and efficiency of communication with health care insurers. Automated clinical summaries reduce clinician effort but also improve consistency of communication and understanding of RT terms for those outside of radiation oncology.

The creation of standardized O3 aims to have a direct effect on gathering quality information from databases across medical systems. Standardization of the nuances of RT delivery across facilities can not only allow for reliable pooling of data between centers but could be integrated into national EHR consistency. By having standard data elements, EHR systems could extrapolate and integrate data within and across facilities as well as to insurers and payers. Once within the EHR in a standardized format, it can also help facilitate automatic documentation, improving efficiency within the clinics. The ontology can be used to inform categorization within Research Electronic Data Capture (REDCap), a secure online web application for databases, which allows for collaborative research projects not currently possible due to the various differences in volume naming, radiation prescriptions, and so on, currently in practices in the United States.

The O3 has been leveraged in the development of HL7 FHIR tags in the CodeX community. For example, the RTTD group started with the completion note use case. It was anticipated that effort would take longer than it has. The work with O3 and multi-institutional engagement greatly accelerated RTTD development. O3 has expanded beyond the first task for RTTD. The RTTD use case team is now working to extend including items identified as high priority in O3 while coordinating with the larger community of CodeX users beyond RT.

Relational databases are the backbone of enterprise level information systems. The relationships identified in O3 lend themselves to the subject-predicate-object form of identifying relationships that are used in triple store databases. By design, the committee included members who were also primary leaders in design and implementation of triple store databases and OBO foundry-based ontologies. These members are also part of a working group operating under BDSC to create a sematic web ontology. Multistakeholder collaboration in creating O3 that creates operational standardizations for aggregation and dissemination provides a foundation for subsequent use of the data in triple stores. We anticipate that using the standardizations of O3, including the database schema rendering, will help to bridge collection of data in relational databases to eventual use of large-scale data sets in inferencing applications with triple stores.

We anticipate that quantifying and standardizing the relationships, key elements, attributes, and value sets will lay a foundation for subsequent development of automated AI algorithms to traverse O3 data sets to promote new discoveries in real-world EHR data. If the standards become adopted widely, then they can enable data from multiple research studies to be combined. For example, it then becomes possible to create algorithms that can automatically investigate the combined effects of chemotherapy, RT treatment, and dosimetric details on outcomes and toxicities from multiple research studies if they use the O3 standard. Subsequent professional society guided refinement of O3 and the SQL rendering will be informed by experience with implementations supporting clinical practice quality improvement and research.

Health care domain-specific knowledge is needed to develop standardizations that can be implemented at point of entry to improve data fidelity. The range of ostensible data "standards" needed to support a health care specialty often fall far short of their promise and practicability, when developed and imposed from outside of the domain expertise of clinicians. For example, Kush et al[49] recently highlighted disappointing results after creation of common data element sets (CDEs) developed as part of siloed NIH efforts. They found that "despite the promise and promulgation of CDEs over the past two decades, most are essentially a local resource and are not suitable for wholesale adoption and global reuse."[49] The O3 provides support for implementing the principles of findable, accessible, interoperable, and reusable (FAIR) data and with the goal to break down these silos.[50] By standardizing data elements, attributes, and value sets, along with creating a standardized database schema, the O3 helps enable data sets to be both interoperable and reusable. By adopting the database schema, it can become possible to share queries among researchers and studies, making data more accessible. Its collaborative framework provides a strong foundation for the continued partnership of professional societies in charting common courses in data science and data-driven practice quality initiatives.

BDSC is currently working with professional societies to incrementally add disease site-specific attributes for inclusion of all ASTRO-defined disease site specialties. Updates and extensions to O3 will be published to the public website along with software tools for creating O3-based databases and interfaces. A subgroup of the BDSC is also formalizing the ontology according to OBO Foundry principles, increasing both its exposure and its interoperability with other ontologies.[51]

Ultimately, this work is intended to harvest the data generated during patient care into forms that are more useful for discovery and safety, which will be ultimately beneficial

for future patients. Data that can be validated will allow testing of sophisticated and provocative hypotheses and will move the field forward more quickly. We see this work as helping patients, helping to advance clinical research, and aiding in optimal resource utilizations as we move into a future featuring personalized, adaptive, motion-managed, imaged-guided, and biologically targeted combined-modality therapies.

Recommendations

Over the course of several meetings, members debated primary recommendations for clinics to engage O3 and increase the volume of high priority data available. Discrete, bulleted recommendations for clinics, researchers, professional societies, and manufacturers are summarized in Table 2. Investment in recommendations for manufacturers

Table 2 Recommendations to clinics, manufacturers, granting agencies, and professional societies for leveraging O3 to meet mutual objectives for quality, safety, improved workflow, and FAIR data sets

No.	Recommendations
1	A. For manufacturers of electronic record systems: Radiation oncology information systems and treatment planning systems applications should implement O3 in user interfaces ordered by the prioritized tables. The standardized JSON formatted O3 file can be used to automatically populate drop-down lists in user interfaces, including mappings to other coding systems (SNOMED CT, NCI, etc), and allow periodic updates as O3 is updated. This structure would support interoperability among multivendor systems and design of clinic customizable "forcing functions" based on professional society-based standards to ease data collection as part of routine practice. B. Use the O3 relationships to ensure database recording methods enable ability to reliably extract high-value related items and relationships (eg, from diagnosis to treatment course and outcomes). Currently, even the highest priority elements and relationships that are routinely needed by physicians are not well optimized for entry, retrieval, and connection.
2	A. For RT centers: O3 should be embedded within the program's quality framework that focuses on RT data standardizations. Centers may choose to implement O3 in a stepwise fashion guided by prioritization of clinical relevance and ease of implementation. To start, full implementation of Table 1 (first priority for aggregation efforts) should be achievable within 1-2 y for many clinics and within 3 y for the vast majority. B. Clinics should examine and adjust clinical processes to assure highest priority attributes are captured in existing EHR and ROIS systems as part of routine use in a fashion supporting accurate, automated electronic capture. C. Many data elements in Table 1 (first and second priorities for aggregation efforts) are currently required by accrediting bodies (eg, ACR, APEx, COC, ASCO). We recommend use of O3 as an organizing reference for accreditation bodies to increase consistency and interoperability of quality metrics.
3	A. For researchers: O3 should form the basis of RT outcomes research with inclusion in clinical trials and grant applications. B. Use the attribute tables and standardized value sets in clinical quality and research designs. This reduces effort and the resulting interoperability supports ability for reuse of data sets. Promoting use of the standardized O3 including value sets and mappings to other systems (eg, NCI Thesaurus), supports government-granting agencies for creating large scale interoperable data sets while synchronizing common objectives with professional societies. C. Use the freely available, O3-based database schema as the basis for creating learning health systems in radiation oncology. This increases ability to interoperable share data and enables development of shared tools for data extraction and analysis to reduce development effort by local staff. D. Encourage NIH and NCI to integrate the standardized O3 JSON data file into data curation applications, such as the POSDA, to create standardized curation tools supporting large scale aggregations of clinical data with imaging data. E. Increase support for consolidation standardization efforts among groups developing coding systems, taxonomies, and ontologies.
4	A. For professional societies: Continue support for collaborative leadership and effort among professional societies in defining standards that support mutual objectives for clinical practice and research. B. Engage with government regulatory bodies to promote the role of professional societies in defining standardizations used as the basis for public health and reimbursement. C. Continue and expand engagement with stakeholders outside of radiation oncology's traditional boundaries (eg, CodeX, USCDI) to promote data policies defined by outside groups that are well aligned to the expertise of professional societies. D. Advocate for standards through groups such as the Global Harmonization Group — Global Quality Assurance of Radiation Therapy Clinical Trials.

Abbreviations: ACR = American College of Radiology; APEx = Accreditation Program for Excellence; ASCO = American Society of Clinical Oncology; COC = Commission on Cancer; CodeX = Common Oncology Data Elements eXtensions; DICOM = Digital Imaging and Communications in Medicine; EHR = electronic health records; FAIR = findable, accessible, interoperable, and reusable; JSON = JavaScript Object Notation; NCI = National Cancer Institute; NIH = National Institutes of Health; O3 = Operational Ontology for Oncology; O3 = Operational Ontology for Oncology; POSDA = Pearl Opens Source DICOM Archive; ROIS = radiation oncology information systems; RT = radiation therapy; SNOMED CT = Systemized Nomenclature of Medicine – Clinical Terms; USCDI = United States Core Data for Interoperability.

and for clinics can provide benefit to the field. Support for standardized data and methods could enable automated coding, which would increase compliance and reduce work associated with code capture and validation. It would also provide a framework for integrating quality measures and procedure codes to facilitate pretreatment payor validations and minimize prior authorization burdens.

Recommendations for clinics and manufacturers can increase patient safety and increase transparency of care for patients by improving consistency of communication and enabling automated documentation to reduce clinician effort.

Recommendations for researchers would support efforts by the NIH and NCI to assemble large databases by reducing barriers for aggregation and increasing interoperability of data across data sets. By grounding these data sets in professional society, multistakeholder, consensus- driven standardizations, focus can shift to more valuable uses of data rather than repeatedly resolving differences between groups. This provides a straightforward approach to integrating clinical and imaging data while respecting the domain expertise and evolving approaches of each specialty. This is also responsive to new requirements in the NIH and NCI for sharable data. Integrating the standardization in O3 using the JSON file, also recommended for manufactures, would enable creation of tools for data curation.

Together with recommendations for manufactures and clinics, recommendations for researchers and professional societies will lower the barriers for aggregation of large data sets that could be used in creating large, representative, FAIR data sets supporting the scientific objectives of grant programs. Curation is typically an intensely manual process, but implementation of the proposed recommendations would pave the way toward development of standardized, automated tools to characterize variation in practice and outcomes with reduced effort and cost.

References

1. Mayo CS, Kessler ML, Eisbruch A, et al. The big data effort in radiation oncology: Data mining or data farming? *Adv Radiat Oncol* 2016;1:260-271.
2. American Society for Radiation Oncology (ASTRO). ASTRO responds to CMS Radiation Oncology Model: Implementation delay and more reforms needed. Available at: https://www.astro.org/News-and-Publications/News-and-Media-Center/News-Releases/2020/ASTRO-responds-to-CMS-Radiation-Oncology-Model-Imp. Accessed June 2, 2022.
3. Mayo CS, Moran JM, Bosch W, et al. American Association of Physicists in Medicine Task Group 263: Standardizing nomenclatures in radiation oncology. *Int J Radiat Oncol Biol Phys* 2018;100:1057-1066.
4. Mir R, Kelly SM, Xiao Y, et al. Organ at risk delineation for radiation therapy clinical trials: Global Harmonization Group consensus guidelines. *Radiother Oncol* 2020;150:30-39.
5. Caissie A, Brown E, Olson R, et al. Improving patient outcomes and radiotherapy systems: A pan-Canadian approach to patient-reported outcome use. *Med Phys* 2018;45:e841-e844.
6. Kildea J, Brown E, Letourneau E, et al. Guidance on the use of common nomenclature in Canadian radiation treatment programs. Available at: https://www.cpqr.ca/wp-content/uploads/2021/04/UCN.2021.03.01.pdf. Accessed June 2, 2022.
7. Gruber TR. A translation approach to portable ontology specifications. *Knowledge Acquisition* 1993;5:199-220.
8. Jackson R, Matentzoglu N, Overton JA, et al. OBO Foundry in 2021: Operationalizing open data principles to evaluate ontologies. *Database* 2021;2021:baab069.
9. National Center for Biomedical Ontology. BioPortal. Available at: https://bioportal.bioontology.org/. Accessed November 17, 2022.
10. American Association of Physicists in Medicine (AAPM). AAPM Committee Tree - Big Data Subcommittee (BDS). Available at: https://www.aapm.org/org/structure/default.asp?committee_code=BDS. Accessed May 22, 2022.
11. HITInfrastructure. Final interoperability rule has implications for APIs, FHIR. Available at: https://hitinfrastructure.com/news/final-interoperability-rule-has-implications-for-apis-fhir. Accessed May 22, 2022.
12. Federal Register. 21st Century Cures Act: Interoperability, information blocking, and the ONC Health IT Certification Program. Available at: https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification. Accessed June 2, 2022.
13. Confluence. Radiation Therapy Treatment Data for Cancer. Available at: https://confluence.hl7.org/pages/viewpage.action?pageId=76162628. Accessed May 24, 2022.
14. Katsoulakis E, Kudner R, Chapman C, et al. Consensus quality measures and dose constraints for head and neck cancer with an emphasis on oropharyngeal and laryngeal cancer from the Veterans Affairs Radiation Oncology Quality Surveillance Program and American Society for Radiation Oncology (ASTRO) expert panel. *Pract Radiat Oncol* 2022;12:409-423.
15. National Library of Medicine. SNOMED CT. Available at: https://www.nlm.nih.gov/healthit/snomedct/index.html. Accessed May 22, 2022.
16. Mayo CS, Phillips M, McNutt TR, et al. Treatment data and technical process challenges for practical big data efforts in radiation oncology. *Med Phys* 2018;45:e793-e810.
17. Mayo C. Community science and reaching the promise of big data in health care. *Med Phys* 2018;45:e790-e792.
18. McNutt TR, Bowers M, Cheng Z, et al. Practical data collection and extraction for big data applications in radiotherapy. *Med Phys* 2018;45:e863-e869.
19. Christodouleas JP, Anderson N, Gabriel P, et al. A multidisciplinary consensus recommendation on a synoptic radiation treatment summary: A Commission on Cancer Workgroup report. *Pract Radiat Oncol* 2020;10:389-401.
20. Hayman JA, Dekker A, Feng M, et al. Minimum data elements for radiation oncology: An American Society for Radiation Oncology consensus paper. *Pract Radiat Oncol* 2019;9:395-401.
21. Kapoor R, Sleeman WC, Nalluri JJ, et al. Automated data abstraction for quality surveillance and outcome assessment in radiation oncology. *J Appl Clin Med Phys* 2021;22:177-187.
22. Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;67:401-409.
23. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Med Teach* 2017;39:14-19.
24. Wang M, Pantell MS, Gottlieb LM, Adler-Milstein J. Documentation and review of social determinants of health data in the EHR: Measures and associated insights. *J Am Med Inform Assoc* 2021;28:2608-2616.
25. Mayo C, Feng M, Brock K, et al. American Association of Physicists in Medicine — Operational ontology for radiation oncology. Available at: https://aapmbdsc.azurewebsites.net/Home/About?ReturnUrl=%2F. Accessed February 27, 2022.
26. HL7 International. Extension: Radiotherapy modality and technique extension. Available at: https://hl7.org/fhir/us/mcode/StructureDefinition-mcode-radiotherapy-modality-and-technique.html. Accessed May 24, 2022.

27. HL7 International. Resource profile: Radiotherapy volume profile. Available at: https://hl7.org/fhir/us/mcode/StructureDefinition-mcode-radiotherapy-volume.html. Accessed May 24, 2022.
28. Mohan G, Chattopadhyay S. Cost-effectiveness of leveraging social determinants of health to improve breast, cervical, and colorectal cancer screening: A systematic review. *JAMA Oncol* 2020;6:1434-1444.
29. Alcaraz KI, Wiedt TL, Daniels EC, Yabroff KR, Guerra CE, Wender RC. Understanding and addressing social determinants to advance cancer health equity in the United States: A blueprint for practice, research, and policy. *CA Cancer J Clin* 2020;70:31-46.
30. Chen M, Tan X, Padman R. Social determinants of health in electronic health records and their impact on analysis and risk prediction: A systematic review. *J Am Med Inform Assoc* 2020;27:1764-1773.
31. Kwon IWG, Kim SH, Martin D. Integrating social determinants of health to precision medicine through digital transformation: An exploratory roadmap. *Int J Environ Res Public Health* 2021;18:5018.
32. Centers for Disease Control and Prevention. Disability and health data system (DHDS). Available at: https://www.cdc.gov/ncbddd/disabilityandhealth/dhds/index.html. Accessed May 24, 2022.
33. American Society for Radiation Oncology. Radiation oncologists urge Congress to reverse proposed CMS cuts and create more equity in access to cancer treatments. Available at: https://www.astro.org/News-and-Publications/News-and-Media-Center/News-Releases/2021/Radiation-oncologists-urge-Congress-to-reverse-pro. Accessed May 24, 2022.
34. Office of the National Coordinator for Health Information Technology. United States Core Data for Interoperability (USCDI). Available at: https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi. Accessed May 22, 2022.
35. Barcelona V, Montalvo-Ortiz JL, Wright ML, et al. DNA methylation changes in African American women with a history of preterm birth from the InterGEN study. *BMC Genom Data* 2021;22:30.
36. Das S, Hussey P. ContSOnto: A formal ontology for continuity of care. *Stud Health Technol Inform* 2021;285:82-87.
37. Hu L, Ma J, Cao M, et al. Exosomal mRNA and lncRNA profiles in cord blood of preeclampsia patients. *J Matern Fetal Neonatal Med* 2021;1-11.
38. Haimel M, Pazmandi J, Heredia RJ, et al. Curation and expansion of human phenotype ontology for defined groups of inborn errors of immunity. *J Allergy Clin Immunol* 2021 S0091-6749(21)00732-6.
39. Traverso A, van Soest J, Wee L, Dekker A. The radiation oncology ontology (ROO): Publishing linked data in radiation oncology using semantic web and ontology techniques. *Med Phys* 2018;45:e854-e862.
40. Ju C, Zhang S. Doctor recommendation model based on ontology characteristics and disease text mining perspective. *Biomed Res Int* 2021;2021 7431199.
41. Michalowski M, Wilk S, Michalowski W, et al. A health eLearning ontology and procedural reasoning approach for developing personalized courses to teach patients about their medical condition and treatment. *Int J Environ Res Public Health* 2021;18:7355.
42. Prieto Santamaría L, Fernández Lobón D, Díaz-Honrubia AJ, Ruiz EM, Nifakos S, Rodríguez-González A. Towards the representation of network assets in health care environments using ontologies. *Methods Inf Med* 2021;60:e89-e102.
43. Steiner B, Saalfeld B, Elgert L, Haux R, Wolf KH. OnTARi: An ontology for factors influencing therapy adherence to rehabilitation. *BMC Med Inform Decis Mak* 2021;21:153.
44. Belenkaya R, Gurley MJ, Golozar A, et al. Extending the OMOP common data model and standardized vocabularies to support observational cancer research. *JCO Clin Cancer Inform* 2021;5:12-20.
45. Belenkaya R, Gurley M, Dymshyts D, et al. Standardized observational cancer research using the OMOP CDM oncology module. *Stud Health Technol Inform* 2019;264:1831-1832.
46. Maier C, Kapsner LA, Mate S, Prokosch HU, Kraus S. Patient cohort identification on time series data using the OMOP Common Data Model. *Appl Clin Inform* 2021;12:57-64.
47. Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The usage of OHDSI OMOP - A scoping review. *Stud Health Technol Inform* 2021;283:95-103.
48. Center for Healthcare Improvement and Medical Effectiveness. Observational Medical Outcomes Partnership (OMOP). Available at: https://chime.ucsf.edu/observational-medical-outcomes-partnership-omop. Accessed May 22, 2022.
49. Kush RD, Warzel D, Kush MA, et al. FAIR data sharing: The roles of common data elements and harmonization. *J Biomed Inform* 2020;107 103421.
50. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3 160018.
51. Phillips MH, Serra LM, Dekker A, et al. Ontologies in radiation oncology. *Phys Med* 2020;72:103-113.