

Европейский проект в информатике материалов. Онтологии и виртуальные платформы

Рассмотрен опыт Европейских организаций по созданию масштабной инфраструктуры, обеспечивающей задачи моделирования материалов, непрерывно возникающие при разработке новых технологий. Инфраструктура объединяет информационно-вычислительные ресурсы и многочисленные сервисы, позволяющие применить новую бизнес-модель в сфере высоких технологий. Ведущую роль в организации подобной инфраструктуры играет система онтологий для стандартизации понятийного аппарата, фигурирующего в используемых документах, моделях и программных средствах. Основные идеи при реализации моделирования систематизированы посредством онтологии ЕММО (European Materials Modelling Ontology), которая способна к расширению и охвату новых категорий объектов и методов исследования. Описаны уникальные стандарты, положенные в основу онтологий и покрывающие практически все виды материальных объектов, от атомов и вплоть до строительных материалов.

Ключевые слова: материалы, материаловедение, платформа как бизнес-модель, виртуальный рынок, данные, интенсивное использование данных, онтологии, стандарты, поток работ

DOI: 10.36535/0548-0027-2021-11-2

ВВЕДЕНИЕ

В эпоху интенсивного вторжения в науку технологий работы с данными требуется интеграция навыков, методологий и данных из множества дисциплин. В этом случае необходимо создание инфраструктур, обеспечивающих взаимосвязь информационных ресурсов и условий работы с ними различных категорий пользователей. Материаловедение служит идеальным полигоном для отработки новых стандартов организации науки, прежде всего, за счет исключительного разнообразия материалов, их особенностей и объема данных, непрерывно поступающих из лабораторий и производства. Как следствие, в науке о материалах на ведущую роль выдвинулась технология цифровых данных, ориентированная на задачи их хранения, обработки и распространения. Ключевым моментом, отличающим новый этап в развитии *materials informatics*, является создание, взамен отдельных ресурсов, масштабной инфраструктуры, объединяющей под единым управлением репозитории данных, сервисы доступа к ним, средства анализа и визуализации.

В настоящей статье изучены возможности подобных инфраструктур, основываясь на опыте Европейского Сообщества в реализации проектов моделирова-

ния вещества и создания специализированных информационно-вычислительных средств. По-видимому, ключевым элементом в этом опыте является перенос в науку практики использования так называемых «многосторонних платформ» (*multi-sided platform*). Сразу же заметим, что это понятие радикально отличается от принятого в компьютерном мире понимания платформы, как «операционной системы или комбинации аппаратных средств и операционной системы» [1]. Речь идет о качественно новой бизнес-модели, выстроенной в мировой экономике вокруг задачи извлечения особого типа сырья – «данных» [2, 3]. Будучи полностью основанной на высоких технологиях, эта модель предлагает участникам среду для обмена информацией, продуктами или сервисами. За счет множества преимуществ, модель охватила самые разные сферы жизнедеятельности: торговля, финансы, транспорт, сетевой сервис и др.

Здесь, в науке о материалах, роль платформы состоит в том, чтобы на общем плацдарме свести представителей индустрии с учеными, способными предложить разработку моделей, программного обеспечения, а также сервисы по обучению, сертификации продукции, разработке стандартов и проч. Помимо того что платформа напрямую сводит заказчика и исполнителя заказа, она реализует для участников рынка тех-

нологии анализа и визуализации данных, моделирования рабочих процессов, облачных вычислений, а также протоколы и стандарты взаимодействия участников. Все участники такого рынка получают выгоды и преимущества, недоступные при традиционных формах постановки исследований.

Эффективность платформы проявляется не только в предоставлении ресурсов и сервисов, но и в поддержке технологических и организационных стандартов деятельности. Последние включают требования к интероперабельности (совместимости программно-аппаратного обеспечения), описанию и сертификации продукции, правилам лицензирования, заключения контрактов и т.п.

На сегодняшний день в материаловедении возникли стандарты, способные унифицировать моделирование материалов, задавая классификацию всего доступного множества физических моделей, а также требования к их спецификации и использованию. Стандарты важны для того, чтобы согласовывать в пределах единой платформы описание данных, моделей и рабочих процессов. Формализовать принятые стандарты и реализовать интероперабельность как способность к обмену данными без вмешательства человека невозможно в отсутствие онтологий. Онтология обеспечивает семантически точное и машинно-обрабатываемое определение сущностей и их соотношений. Она включает таксономически организованный словарь понятий в сочетании с логическими соотношениями и аксиомами, формализующими содержание предметной области [4, 5]. Богатый потенциал онтологий, позволил внедрить различные версии предметно-ориентированных онтологий для медицины, химии, нанотехнологий и других дисциплин. Традиционно задача предметно-ориентированной онтологии состояла в выделении базовых сущностей, например, веществ и реакций в химии или категорий материалов и их свойств в материаловедении [6]. Иначе, онтология фокусируется на методах и организации исследований, претендуя на роль управляющей надстройки в *Data-Intensive Science*.

При переходе к «платформенной» организации исследований разработка и применение онтологий заняли одно из ведущих мест в реализации задач виртуального рынка: подбора моделей для индустрии, обмена данными по свойствам и условиям эксплуатации материалов, технологиям обработки данных и т.п. При этом они решают ключевую задачу по семантической *интероперабельности* цифровой платформы, что позволяет программам работать с различными ресурсами данных и добавлять “понимание” (семантику) к данным, продвигая их на уровень информации [7].

МНОГОСТОРОННИЕ ПЛАТФОРМЫ И ИХ ПОТЕНЦИАЛ В НАУКЕ

Несмотря на отсутствие в литературе четкой дефиниции, большинство авторов подчеркивают, что многосторонняя платформа это именно бизнес-модель, предусматривающая торговую площадку для нескольких участников рынка, с предоставлением для них разнообразных ресурсов и поддержкой технологических

и управленческих стандартов [2, 3, 8-10]. Обеспечив наиболее эффективные способы извлечения и использования данных, модель распространилась по всей экономике, затронув как технологические компании (Google, Amazon), так и промышленные гиганты (General Electric, Siemens). Выделяют платформу именно как бизнес-модель следующие характерные признаки:

- коммерческая активность в определенной сфере;
- открытость рынка для новых участников с предоставлением для них всего потенциала в виде ресурсов и сервисов;
- использование цифровой платформы для хранения и обработки данных и сетевых коммуникаций;
- единая система стандартов, документации, правил заключения контрактов и т.п.

При этом, используя платформу, участники рынка получают большую прибыль, чем при традиционных взаимодействиях [2, 9]. Это получило название *сетевого эффекта*, когда ценность платформы для отдельного участника возрастает с их общим числом. При этой модели возникает и косвенный сетевой эффект, когда потребление продукта или услуги влечет дополнительное использование других (продуктов, услуг), что, в конечном счете, сказывается и на потреблении исходного продукта или услуги.

Наряду с традиционными для подобных платформ видами деятельности (коммуникации, инвестиции, социальные сети), эта модель проникла и в сферу инноваций. Нарботанные стратегии оказались востребованы в науке, прежде всего там, где прогресс неразрывно связан с привлечением и использованием обширных массивов данных. Этот подход позволил радикально изменить типовую практику материаловедения, когда на смену длительным и дорогостоящим испытаниям пришли цифровые технологии, объединяющие систематизацию данных и способы их анализа на основе методов машинного обучения. Появились многочисленные инфраструктуры, которые в рамках одной среды объединяют базы данных (БД), программы компьютерного моделирования и анализа данных (их перечень в области материаловедения можно найти в обзоре [11]).

Новый элемент, внесенный в эту деятельность Европейским Советом по моделированию материалов (The European Materials Modelling Council – ЕММС), это соединение научного потенциала цифровых инфраструктур с бизнес-моделью. Возможности цифровых технологий тогда расширяются за счет:

- создания обширной социальной среды из экспертов разного профиля;
- структурирования среды с выделением участников, различающихся по видам спроса и предложений на услуги из области материаловедения;
- подготовки и принятия участниками рынка базовых документов по систематизации материалов и средств моделирования: словарей, классификаторов и онтологий.

Подробный план исследований и мероприятий, проводимых ЕММС, можно найти на сайте [12]. Две категории участников играют ведущую роль, по существу определяя главную задачу и рыночный потенциал

платформы: представители промышленности как заказчики и ученые-материаловеды, способные реализовать поставленные задачи. Заключение и поддержка контрактов между этими двумя группами участников обеспечивают, в основном, рыночную основу функционирования платформы.

Однако для реализации поставленных промышленностью задач, учитывая их сложность и многообразие, следует привлечь много других участников. Весьма специфична роль экспертов, знающих особенности производства и помогающих как в подборе физических моделей, способных описать свойства материала при его эксплуатации, так и в подборе специалистов, способных предложить требуемую модель и соответствующее программное обеспечение (ПО). По существу, их роль состоит в переводе поставленной задачи с языка, принятого в инженерной среде, на язык, принятый в науке о материалах¹. Именно эта категория участников обеспечивает заключение контрактов и выигрыш, получаемый за счет постановки моделирования.

Поскольку любая из физических моделей (например, кинетика коррозии или теория прочности) реализуется на компьютере, поставка стандартных или разработка оригинальных кодов, а также их отладка, тестирование и сопровождение требуют привлечения владельцев и разработчиков ПО. Аналогично работа с данными предусматривает присутствие на рынке владельцев или администраторов БД, а также специалистов по интеллектуальному анализу, прежде всего по методам машинного обучения.

Наконец, функционирование платформы обеспечено наличием многочисленных онтологий и словарей, документирующих все множество возможных моделей, методов решения, постановок задач и т.п. Их создание требует привлечения специалистов, сочетающих знания в предметной области с опытом систематизации данных. Координированная работа в рамках платформы разных категорий участников позволяет каждой из этих категорий добиться конкурентных преимуществ, недоступных при автономной работе. В табл. 1 приведен их перечень, отмеченный в докладе [3].

Как показано в публикациях семинара по проблемам интероперабельности ресурсов [7], конкурентные преимущества и добавленная стоимость возникают также за счет извлечения знаний из обширных массивов данных. Анализ данных по свойствам доступных материалов позволяет оценить риски и расходы, связанные с разработкой новых материалов, совершенствованием конструкций и технологий, и как следствие, инициировать новые заказы для специалистов по моделированию.

Важные преимущества обеспечивает распространение принятых в рамках платформы стандартов и онтологий, обеспечивающих интероперабельность при связывании разнородных данных. Богатый потенциал онтологий в работе платформы, обеспечивающей науку о материалах, дает организация Вир-

туального рынка (The Virtual Materials Marketplace – VIMMP) [13], созданного Европейским Советом. Чтобы обеспечить полностью интегрированную среду, VIMMP строится на базе стандартов и онтологий, которые сосредоточены на моделях, программных инструментах, экспертных знаниях по переводу и учебных материалах. Не ограничиваясь решением научных проблем (систематизация моделей и характеристик материалов), создатели рынка формализовали посредством онтологий, по сути, все аспекты в жизнедеятельности платформы [14].

В рамках виртуального рынка VIMMP реализуется набор взаимно-связанных онтологий, перечень которых с анализом функций приведен в публикациях [14, 15]. В табл. 2 приведен сокращенный перечень из 8 «рыночно-ориентированных» онтологий с пояснением их возможностей в поддержке работы рынка. Первая группа (№1–4) стандартизует вспомогательную деятельность, например работу интерпретатора (подбор моделей для заказчика) или коммуникации (онтология VIVO). При этом, кодируя на специальном языке определенный набор концепций и видов активности, онтология часто опирается на ранее принятый текстовый стандарт, скажем, руководство интерпретатора [16] или на классификаторы, созданные Ассоциацией вычислительной техники (www.acm.org/publications/class-2012/) и Американским физическим обществом (physh.aps.org).

Вторая группа онтологий (№5–8) ответственна за семантическую интеграцию концепций и лексики, относящихся к научно-технологическому сегменту: модели (№5), программно-аппаратные средства (№6,7), физические свойства материалов (№8). Особая роль принадлежит онтологии OSMO (№5), которая формализует стандарт MODA, решающий исключительную по сложности задачу: классификацию и описание моделей практически всех видов, а также регламент рабочих процедур, сопровождающих постановку моделирования.

На рис. 1 из отчета [14] показана взаимосвязь онтологий рынка VIMMP, каждая из которых вынуждена заимствовать концепции из других для адекватного представления своего сегмента предметной области. Так, онтология OSMO, отражающая проблемы моделирования, очевидным образом связана с онтологией VISO, отражающей тематику ПО, и онтологией MMTO, отражающей концепции, обеспечивающие работу интерпретаторов при подборе моделей по заказу промышленности.

Заметим, что в сравнении с табл. 2, список онтологий на рис. 1 расширен на две: ЕММО (Европейская онтология моделирования материалов) и подчиненная ей ЕVMPO (онтология Европейской виртуальной торговой площадки). Первая из них, описанная далее, суммирует именно научные концепции, привлекаемые при моделировании, отвлекаясь от понятийного багажа, связанного с функционированием рынка (инфраструктура и сервисы). Будучи онтологией, так называемого, верхнего уровня, ЕММО имеет семантическую совместимость с онтологиями рынка, перечисленными в табл. 1. Эта связь обеспечена обязательным использованием во всей документации и информационных

¹ Для этой категории специалистов в англоязычной литературе используется термин *translator*.

потоках набора из 11 фундаментальных категорий: (1) оценка (assessment), (2) календарное событие (calendar event), (3) коммуникация (communication), (4) информационная сущность (information content entity), (5) инфраструктура (infrastructure), (6) интерпретатор (interpreter), (7) материал (material), (8) модель (model), (9) процесс (process), (10) продукт (product), (11) свойство (property). Эти категории связаны с соответствующими сущностями из ЕММО через модуль интеграции ЕММО-VIMMP.

Таблица 1

Возможности, предоставляемые платформой для участников рынка

Участник рынка	Преимущества использования платформы
Создатели моделей	Наличие дополнительных моделей и ПО, поиск заказчиков и коллективов для сотрудничества
Малые предприятия	Новые знания, доступность экспертизы
Исследовательские лаборатории на предприятиях	Дополнительные ресурсы, возможности использования ПО и постановки моделирования
Поставщики и разработчики ПО	Новые клиенты и сотрудники, тестирование продукта
Интерпретаторы (<i>трансляторы</i>)	Новые заказчики, знания, доступ к специалистам и инструментам моделирования
Лекторы и консультанты	Возможности демонстрационного использования средств моделирования

Таблица 2

Онтологии рыночного уровня

№ п/п	Онтология	Предметная область/сфера применения
1	The Materials Modelling Translation Ontology (MMTO)	Онтология деятельности интерпретатора по выбору адекватного процесса моделирования для конкретных заказов от промышленности. Построена путем формализации руководства интерпретатора [16]
2	The VIMMP Communication Ontology (VICO)	Онтология коммуникаций охватывает метаданные сообщений, которыми обмениваются участники виртуального рынка: пользователи, разработчики моделей, владельцы софта и т.д. Онтология опирается на стандарт ISO 3166 для обозначения стран и регионов и софта
3	The Ontology for Training Services (OTRAS)	Онтология обучающих сервисов. Формализует описания учебных курсов, программ, результатов обучения и т.п. Включает таксономию, допускающую использование классификаторов Ассоциации вычислительной техники и Американского физического общества
4	The VIMMP Validation Ontology (VIVO)	Онтология контроля достоверности, включает метаданные, отражающие ошибки моделей, решателей и процессов обработки данных, оценки требований к вычислительным ресурсам, и сопоставительному анализу. Онтология отражает также обратную связь с клиентами на платформе виртуального рынка
5	The Ontology for Simulation, Modelling, and Optimization (OSMO)	Онтологическая версия полужормального текстового и графического стандарта MODA
6	The Marketplace-Accessible Computational Resource Ontology (MACRO)	Вычислительные и аппаратные ресурсы, включая форматы файлов
7	The VIMMP Software Ontology (VISO)	Онтология программного обеспечения отражает сведения о предлагаемых на рынке программных средствах, их возможностях, в том числе аспекты лицензирования. Включает три модуля, реализующие отдельные категории моделей: электронные (viso-el), атомистические+мезоскопические (viso-am) и континуальные (viso-co). Ссылается на ранее существовавшую онтологию программного обеспечения The Software ontology (SWO) ¹
8	The VIMMP Ontology of Variables (VOV)	Охватывает физические свойства, используемые моделями и решателями. Непосредственно связана с онтологиями VISO и OSMO. Использует единицы измерения, принятые в онтологии QUDT ²

ПРИМЕЧАНИЯ: ¹ Software ontology (www.obofoundry.org/ontology/swo.html).

² Quantities, Units, Dimensions, and Types Ontology (QUDT, www.qudt.org).

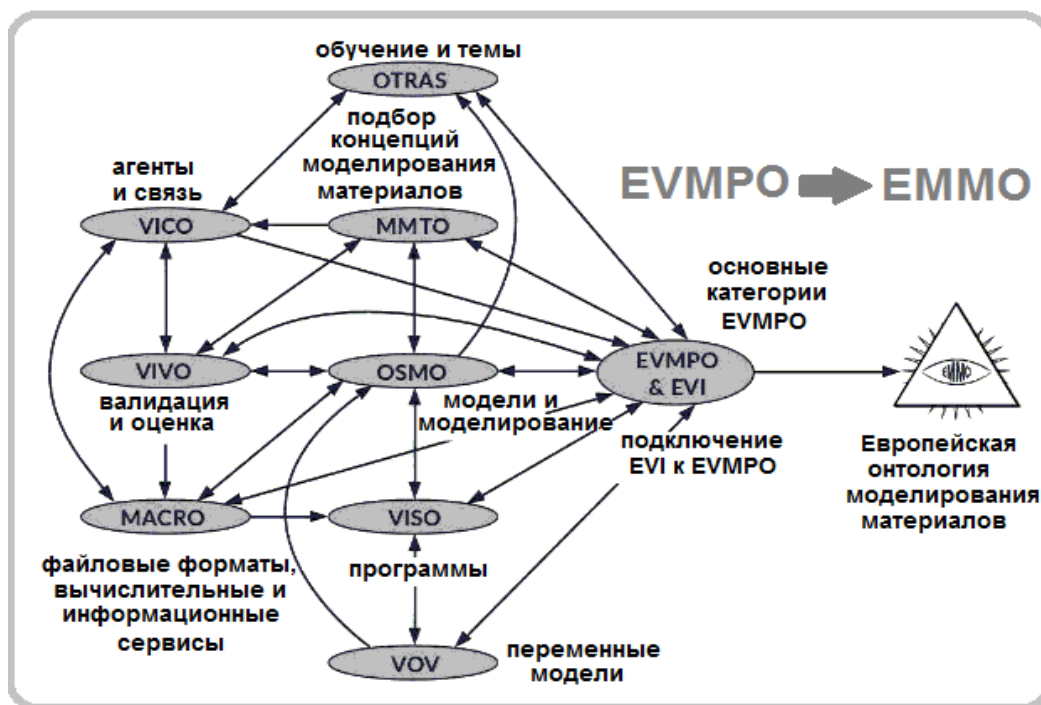


Рис. 1. Онтологии, отражающие функции виртуального рынка. Стрелки указывают связь между концепциями разных онтологий [14].

Указанная совокупность онтологий строго формализует и унифицирует научные концепции, а также все, что относится к бизнес-процессам в пределах VIMMP, своеобразного рынка, предлагающего продукты и услуги, связанные с моделированием материалов. Однако онтологии кодируют на специальном языке все требования и понятия, которые до этого были сведены в текстовом стандарте MODA.

СТАНДАРТ MODA (MOdelling DATA)

Базовым элементом этого стандарта является набор шаблонов (анкет), заполнение которых дает пошаговое документирование информации, отражающей принципы и реализацию модели². Документ [17], определяет словарь и метаданные для заполнения шаблонов, а руководство [18] предлагает каталог, включающий множество физически обоснованных моделей с детализацией уравнений, методов решения и представления результатов. Записанный в виде простого текста стандарт (в отличие от онтологий) не требует для восприятия и использования каких-либо компьютерных средств.

Все богатство и многообразие проблем, связанных с постановкой моделирования, сведено в стандарте к детальному раскрытию в шаблонах всего четырех тем:

1. Прецедент использования (*use case*) и/или описание изучаемой системы (материал, геометрия, условия моделирования).

2. Модель, определяемая выбором физических уравнений и материальных соотношений.

3. Решатель, включая выбор вычислительных средств, численного метода, ПО, временного шага и других параметров численного метода.

4. Средства и процедуры обработки данных на выходе вычислительного процесса (метод, усреднение данных, оценка неопределенности).

При заполнении шаблонов пользователь следует записанной там инструкции о содержании ответов на детализирующие вопросы, представляя их в виде свободного текста произвольного объема. Всего выполняются 5 типовых документов:

- 1) обзор моделирования (вводный документ);
- 2) детализация прецедента (*aspects of the user case*);
- 3) уравнения модели;
- 4) численный метод и ПО;
- 5) обработка результатов.

В качестве примера в табл. 3 представлен шаблон первого из этих документов с инструкциями по представлению модели. Рубрика *Chain of Models* (набор моделей) в первом документе предусматривает возможность использования нескольких моделей, привлекаемых для решения конкретной задачи. При этом в документе четко различаются модели двух базовых категорий: физически обоснованные и, так называемые, *data-based models* (т. е. использующие методы статистического анализа, машинного обучения и проч.). К физически обоснованным стандарт относит все виды моделей, использующих уравнения из разных отделов физики, вне зависимости от уровня их строгости. Так, наряду с методами квантовой статистики для многоэлектронных систем, возможно привлечение и упрощенных инженерных формул для оценки прочности или теплообмена.

² Все шаблоны можно скачать с сайта <https://emmc.info/moda/>

**Шаблон первого/вводного документа при стандартизованном описании
проекта моделирования материала (<https://emmc.info/moda/>)**

ОБЗОР МОДЕЛИРОВАНИЯ							
1	<p>СЦЕНАРИЙ ИСПОЛЬЗОВАНИЯ</p> <p><i>Общее описание. Пожалуйста, укажите свойства и поведение конкретного материала, производственный процесс и / или поведение в процессе эксплуатации, которые будут моделироваться.</i></p> <p><i>Никакой информации о моделировании здесь не должно быть. Идея состоит в том, что этот сценарий также может быть смоделирован другими исполнителями с использованием других моделей с последующим сравнением результатов.</i></p>						
2	<table border="1"> <tr> <td align="center">МОДЕЛЬ 1</td> <td> <p><i>Укажите первую модель. Обратите внимание, что модель должна быть из класса физически-обоснованных (physics-based) моделей.</i></p> <p><i>Большинство проектов моделирования состоят из цепочки моделей (поток работ). Здесь должны быть указаны только физические уравнения и использованы названия из руководства [18]. Все модели следует идентифицировать как электронные, атомистические, мезоскопические или континуальные.</i></p> </td> </tr> <tr> <td align="center">МОДЕЛЬ 2</td> <td> <p><i>Укажите вторую модель.</i></p> </td> </tr> <tr> <td align="center">МОДЕЛЬ НА ОСНОВЕ ДАННЫХ</td> <td> <p><i>Если используются модели, основанные на данных, просьба указать.</i></p> </td> </tr> </table>	МОДЕЛЬ 1	<p><i>Укажите первую модель. Обратите внимание, что модель должна быть из класса физически-обоснованных (physics-based) моделей.</i></p> <p><i>Большинство проектов моделирования состоят из цепочки моделей (поток работ). Здесь должны быть указаны только физические уравнения и использованы названия из руководства [18]. Все модели следует идентифицировать как электронные, атомистические, мезоскопические или континуальные.</i></p>	МОДЕЛЬ 2	<p><i>Укажите вторую модель.</i></p>	МОДЕЛЬ НА ОСНОВЕ ДАННЫХ	<p><i>Если используются модели, основанные на данных, просьба указать.</i></p>
МОДЕЛЬ 1	<p><i>Укажите первую модель. Обратите внимание, что модель должна быть из класса физически-обоснованных (physics-based) моделей.</i></p> <p><i>Большинство проектов моделирования состоят из цепочки моделей (поток работ). Здесь должны быть указаны только физические уравнения и использованы названия из руководства [18]. Все модели следует идентифицировать как электронные, атомистические, мезоскопические или континуальные.</i></p>						
МОДЕЛЬ 2	<p><i>Укажите вторую модель.</i></p>						
МОДЕЛЬ НА ОСНОВЕ ДАННЫХ	<p><i>Если используются модели, основанные на данных, просьба указать.</i></p>						
3	<p>ПУБЛИКАЦИЯ С КРИТИЧЕСКИМ АНАЛИЗОМ ДАННЫХ</p> <p><i>Пожалуйста, укажите публикацию, которая документирует данные конкретного вида моделирования. В ней должно быть гарантировано качество набора данных, а не только моделей.</i></p>						
4	<p>УСЛОВИЯ ДОСТУПА</p> <p><i>Пожалуйста, укажите, являются ли модель и / или данные бесплатными, коммерческими или открытыми. Пожалуйста, укажите владельца и название программного обеспечения или базы данных (включите web-ссылку, если таковая имеется).</i></p>						
5	<p>ПОТОК РАБОТ И ЕГО ОБОСНОВАНИЕ</p> <p><i>Пожалуйста, дайте текстовое обоснование того, почему Вы, как разработчик моделей, выбрали эти модели и этот поток работ, зная, что другие разработчики моделей будут моделировать один и тот же случай для конечного пользователя по-разному.</i></p> <p><i>Это должно включать причину, по которой конкретный аспект сценария использования должен быть смоделирован с помощью конкретной модели.</i></p>						

В то же время модели, основанные на данных, базируются на доступных массивах экспериментальных данных и применимы в тех случаях, когда физическая картина не вполне ясна – например, при оценке усталости металла, коррозии, токсичности материала и проч. Модели 1 и 2, указанные в табл. 3, относятся к категории физически обоснованных моделей, в то время как для второй категории стандарт MODA предлагает аналогичный набор из трех документов: детализация прецедента (*aspects of the user case*); *data-based model*; вычислительные детали. Второй из них (*data-based model*) определяет не только метод анализа, но и базу данных, содержащую массив, использованный при обработке.

Все физически обоснованные модели, согласно стандарту, могут быть отнесены к одной из четырех базовых категорий: электронные, атомистические, мезоскопические и континуальные. К мезоскопическим отнесены модели, трактующие свойства молекул, их агрегатов, наночастиц и т.п., а к континуаль-

ным – модели, в которых материал рассматривается на феноменологическом уровне как макроскопический объект (механика сплошной среды, электродинамика, теория тепломассообмена и т.п.). Кроме того, при декларативном описании модели выделяются физические уравнения (Physics Equation, PE) и так называемые материальные соотношения (Materials Relations, MR), определяющие свойства конкретного вещества, например плотность, вязкость и т.д. Строгая и полная дефиниция модели сводится к определению: **Model=Governing Equations=PE+MR.**

Указанные категории дают лишь верхний уровень категоризации. Для охвата реальных ситуаций стандарт предлагает более глубокую детализацию, основанную на классификационной схеме для физических моделей. В табл. 4 показана упрощенная версия такой схемы, включающая 25 моделей, распределенных по четырем указанным категориям [14].

В самом документе [18] эта схема представлена в заметно расширенном виде, включающем несколько

уровней иерархии. Например, раздел «Уравнения движения (атомистические)» в иерархии представлен в три уровня, см. табл. 5.

Принятая формализация позволила систематизировать безграничное множество физических моделей, сценариев и математических средств посредством ограниченного множества базовых концепций. Должны быть заданы четыре сущности, аттестующие модель (прецедент использования, модель, решатель, обработка данных), четыре уровня гранулирования/детализации (электронный, атомарный, мезоскопический, континуальный) и 25 типовых физических моделей (табл. 4). При необходимости перечень моделей расширяют, используя обширный каталог из руководства [18].

Для каждого класса моделей в каталоге приведен очерк с ее подробной спецификацией, т.е. указанием базовых уравнений и сферы применения. Для иллюстрации возможностей стандарта в руководстве [18] дан список из 36 документированных проектов, которые охватывали самые разные отрасли и виды материалов. Например, проект *ALLIANCE: Crash-worthiness*

of car wheel-house направлен на повышение ударной стойкости автомобильной кабины с моделированием циклической усталости материала, вибрационных и прочих характеристик, определяющих аварийную устойчивость. Другой проект *MOSTOPHOS: Stability of organic light emitting diodes* имел целью повышение эффективности и времени жизни светоизлучающих диодов. В этом случае для решения проблем привлекаются модели вакуумного осаждения слоев при изготовлении, процессов триплет-триплетной аннигиляции, отвечающих за эффективность преобразования энергии, а также макроскопических эффектов, отвечающих за диссипацию тепла. Для каждого из проектов, документированных в [18], приведены образцы заполненных шаблонов. Охват множества объектов и процессов, начиная от механики удара и вплоть до процессов на электронном уровне в оптических устройствах, адекватно отраженный при заполнении шаблонов, демонстрирует возможности стандарта для решения индустриальных задач.

Таблица 4

Классификатор физических моделей [14]

Уровень гранулированности	Тип физической модели	
Электронный	<i>Ab-initio</i> методы квантовой механики	Статистическая теория транспорта заряда
	Модели с эффективным гамильтонианом	Статистическая теория спинового транспорта
	Квантово-механическое моделирование нестационарных процессов	
Атомистический	Классическая версия функционала плотности (DFT, атомистическая)	Статистические суммы (атомистические)
	Молекулярная статика (атомистическая)	Атомистические спиновые модели
	Уравнения движения (атомистические)	Статистическая теория переноса (атомистическая)
Мезоскопический	Классическая версия функционала плотности (DFT, мезоскопическая)	Статистические суммы (мезоскопические)
	Молекулярная статика (мезоскопическая)	Мезоскопические спиновые модели
	Уравнения движения (мезоскопические)	Статистическая теория переноса (мезоскопическая)
Континуальный	Механика твердого тела	Термодинамика континуума
	Механика жидкости	Кинетика химических реакций
	Теплоперенос, термомеханика	Электромагнетизм
	Модели фазового поля, DGT	Процессы и приборы

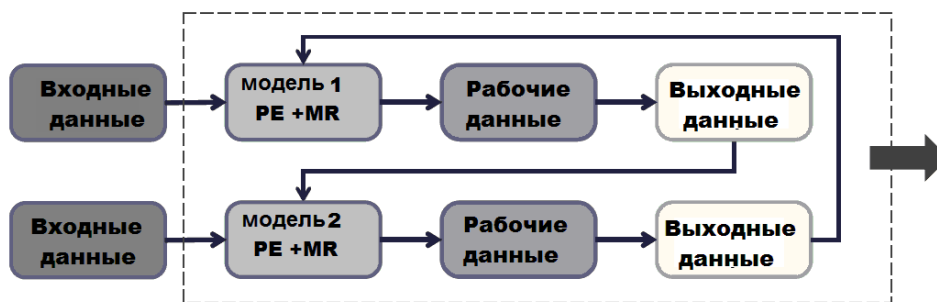
ПРИМЕЧАНИЕ: DFT – density functional theory, DGT – density gradient theory

Таблица 5

Фрагмент иерархии физических моделей [18]

Модели, основанные на уравнениях Ньютона	Молекулярная динамика	Квантовая механика (включена в поток работ)
		<i>Ab initio</i> молекулярная динамика
		Классическая молекулярная динамика
	Молекулярная механика	

Процесс для итеративной связи моделей



Процесс для сильно связанных моделей

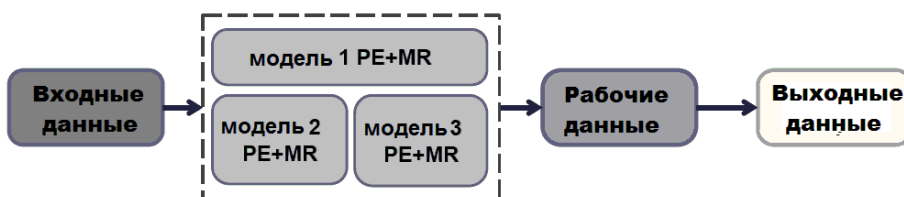


Рис. 2. Образец графической нотации для случаев итеративной и сильной связи двух моделей [20]

Наряду с каталогизацией и описанием моделей, в рамках стандарта MODA регламентируется согласованная работа нескольких моделей и кодов, привлекаемых в процессе моделирования, – так называемый *workflow* (поток работ). Под этим термином понимается координируемое выполнение совокупности взаимосвязанных единиц работы (задач) [1]. Детализация потока работ позволяет формализовать последовательность связанных действий в бизнес-процессах, функционировании механизмов или программно-аппаратных средств, в любом случае представляя абстракцию реальной работы, выделенной в процессе деятельности. Для моделирования потоков работ существуют формальные описания и графические нотации, составленные из операций (работ), логических символов («и», «или») в разветвлениях графической записи, стрелок для отображения последовательности операций или потока ресурсов. Основная сфера применения потока работ – описание и моделирование бизнес-процессов, для чего создан ряд международных стандартов, например BPMN (*Business Process Model and Notation*) [19].

Стандарт MODA, наряду с текстовыми документами, содержит упрощенную нотацию, схематично представляющую потоки работ при согласованном использовании нескольких моделей. Анализ конкретных проектов, приведенных в [18], позволил выявить несколько характерных сценариев. Прежде всего, это последовательная работа (*linking*), когда результаты использования одной модели служат входными данными для последующей. Сценарий с образованием замкнутой петли, т. е. повторной передачей выходных данных второй модели на вход первой, получил название «итеративной связи», реали-

зующей процесс до достижения сходимости. Наконец, часто используется параллельная работа моделей при обеспечении сильной связи (*tight coupling*), когда базовые соотношения в обеих моделях (PE, MR) рассматриваются как единая система уравнений. Для каждой из схем на рис. 2 [20] предложена упрощенная графическая запись с пояснением типа моделей и их связей.

Нотация, включенная в описание проектов в [18], иллюстрирует потоки работ при, так называемом, *многомасштабном* моделировании [21]. Например, моделирование железобетона (проект *LORCENIS: Reinforced concrete*), определяющее на выходе трещиностойкость, старение материала и др. эксплуатационные характеристики, включает работу четверки моделей, начиная с традиционных квантово-механических расчетов для определения коэффициентов диффузии и констант скоростей химических реакций.

ЕВРОПЕЙСКАЯ ОНТОЛОГИЯ МОДЕЛИРОВАНИЯ МАТЕРИАЛОВ (ЕММО)

Если все онтологии рынка VIMMP обеспечивали, в основном, организационные аспекты деятельности, то ЕММО выполняет роль идейного базиса, выполняя семантическую интеграцию данных, моделей и программ, моделирующих свойства и процессы в материалах. Ее главная особенность – полная универсальность в применимости к возможным классам материалов и выбору моделей и типов данных. Достигается тематическая универсальность за счет: (1) использования стандартов [17, 18] в качестве основы систематизации объектов и методов; (2) разделения фундаментальных концепций и промежуточного уровня

для включения концепций из узких областей (металлы, керамика, коррозия и т.п.); (3) заимствования ряда философских концепций, позволяющих на уровне абстракций трактовать процессы измерения и моделирования для произвольных объектов и на всех шкалах атомно-молекулярной структуры вещества.

Структура и содержание онтологии впервые выстроены на основе концепций *семиотики* – философской дисциплины, изучающей строение и функционирование различных знаковых систем, хранящих и передающих информацию [22]. Помимо закрепленных в знаках культурных феноменов, в поле зрения семиотики находятся естественные и искусственные языки, что позволяет формализовать логические и математические системы, включая программы, алгоритмы и языки-посредники для общения с ними человека.

Использование аксиом, заимствованных из семиотики, позволило роль абстрактных объектов в онтологии передать *семиотическим объектам*, т. е. объектам реального мира (знакам или символам), которые указывают на другие объекты, что подлежит интерпретации некоторым агентом (интерпретатор или программа). Тем самым исключается обращение к объектам, которым нельзя приписать протяженность в пространственно-временном континууме. С позиций семиотики свойство надо трактовать как знак, указывающий на материальный объект, воспринимаемый в процессе наблюдения или измерения. Тогда физическая или математическая модель в онтологии также рассматривается как знак, указывающий на физический объект или процесс.

Верхний уровень онтологии содержит специальную ветвь (рис. 3) с тройкой основных элементов семиотики – интерпретатор (*interpreter*), объект (*object*), знак (*sign*), а также концепцию семиозиса (*semiosis*), процесса интерпретации знака или порождения значения, что сводит сложность физического мира к простым знакам.

Другая система концепций связана с так называемой *мереотопологией*, которая представляет синтез *мереологии* (наука о связи «общего и целого», дисциплина восходящая чуть ли не к Аристотелю [23]) с топологией, изучающей геометрические свойства и их сохранение при деформации. На этой основе в исходной точке онтологии вводится ключевое различие между двумя понятиями: *Item* (связанный набор материальных сущностей) и *Collection* (коллекция не связанных объектов)³. Первое из них становится вершиной иерархии, представляющей в ЕММО способность интерпретатора (человек или компьютерная программа) к восприятию объектов окружающего мира на разных уровнях детализации (*гранулированности*), от макроскопических масштабов и вплоть до электронной структуры вещества. Гранулированность представления свойств обеспечена за счет мереологического соотношения *is direct part of* (прямое включение), которое позволяет при рассмотрении объекта выделить элементы, находящиеся при его гранулирова-

нии на ближайшем уровне иерархии, например атомы по отношению к молекулам. Это соотношение радикально отличается от соотношений вида *is proper parthood of*, которые учитывают элементы на всех уровнях иерархии, например по отношению к молекуле, это атомы, электроны, атомные ядра и т.п. Именно понятие о прямом включении позволяет реализовать описание процессов многомасштабного моделирования с последовательным переходом между разными уровнями строения материала.

Детали использования в ЕММО семиотики и мереологии описаны в лекции д-ра А.Симперлера на семинаре, организованном ЕММС (Европейский Совет по моделированию материалов) в 2019 г. [25]. Краткое описание базовых принципов, заложенных в онтологии, можно найти на портале [26].

На основе указанных абстракций сформирован верхний уровень онтологии (рис. 4) с переходом на более глубоких уровнях иерархии к объектам реального мира и отражением их свойств и методов моделирования. Построение иерархии (см. рис. 4) начинается именно с класса *Item*, трактуемого как связанный набор других сущностей.

Суб-классы (*Quantum, Physical, Void*) вводят, наряду с логико-философскими концепциями, базовые принципы физики. В частности, реальные объекты определены как 4D объекты, т. е. одновременно протяженными в пространстве и времени в соответствии с теорией относительности. Класс *Quantum* вводит мереологические соотношения, чтобы разделить пространственные и временные измерения, так что его экземпляры (*instances*) представляют собой наименьшие из возможных объектов 4D мира. Класс *Physical* порождает обширную иерархию реальных физических объектов, а концепция *Void* открывает путь к описанию квантовых систем, где используется понятие «вакуум».

Иерархия, порождаемая классом *Perspective* (*perspektiva*), открывает так называемый средний уровень (*middle level*) для перехода к концепциям, применимым в узких областях знания. При переходе к ним в онтологии выдерживаются два базовых принципа: *редукционизм* (представление объектов как композиций из объектов следующего уровня) и *плюрализм* (множественность трактовок в соответствии с предпочтениями интерпретатора; например, это может быть выбор между микро- и макроскопическим подходом, или между физическими моделями и основанными на обработке данных).

В соответствии с этими принципами на различных уровнях иерархии представлены классы, отражающие реальные виды объектов, процессов и моделей. Например, на рис. 5 показана иерархия класса *Matter*, отражающая существование как материалов разной природы, так и составляющих их частиц. Аналогичная иерархия на рис. 6 иллюстрирует возможность работы с широким кругом моделей, принятых в онтологии в соответствии со стандартом [18]: четыре типа физических моделей в зависимости от уровня гранулированности вещества и отдельный класс моделей (*data-based*), основанных на обработке данных, в основном методами машинного обучения и другими методами искусственного интеллекта.

³ Уже здесь проявляется отличие ЕММО от традиционной концептуализации, когда верхний уровень начинается с выделения сущностей двух типов: *continuant* (не меняющиеся во времени) и *occurrent* (меняющийся или протяженный во времени) [24].

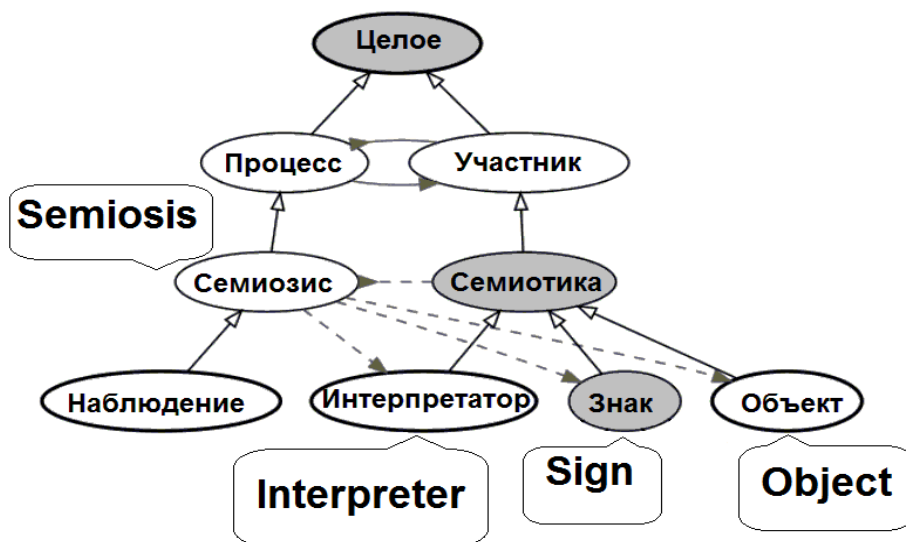


Рис. 3. Ветвь онтологии с концепциями из семиотики (из описания онтологии ЕММО [27])

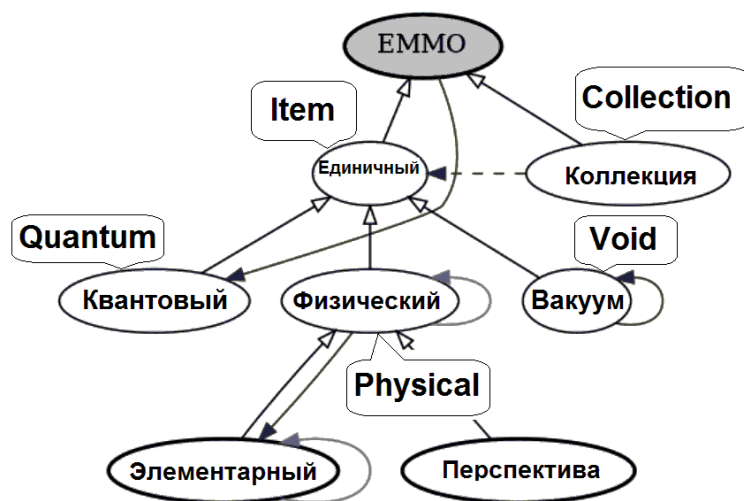


Рис. 4. Верхний уровень онтологии ЕММО (из описания онтологии ЕММО [27])

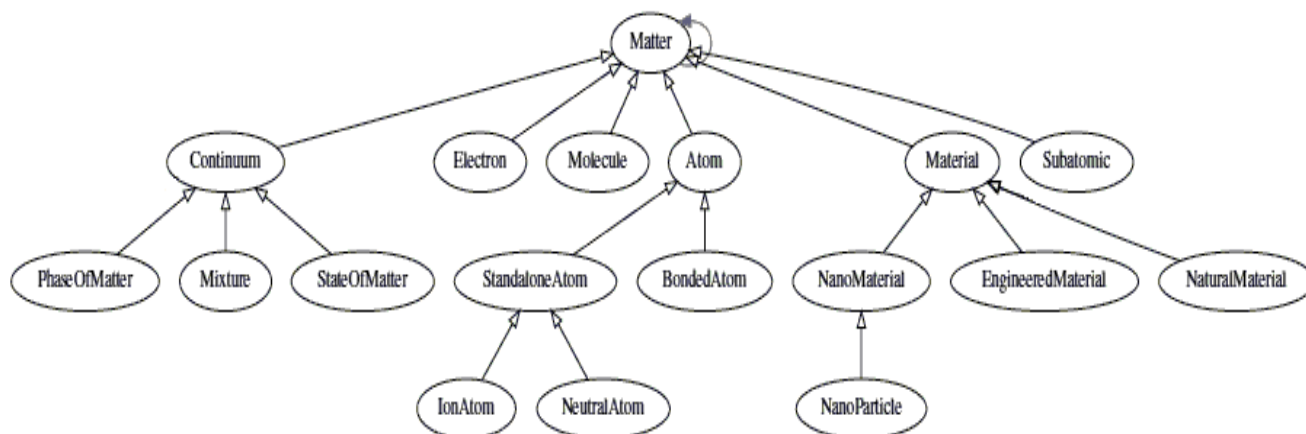


Рис. 5. Matter – одна из ветвей промежуточного уровня, включающая типовые материальные объекты (из описания онтологии ЕММО [27])

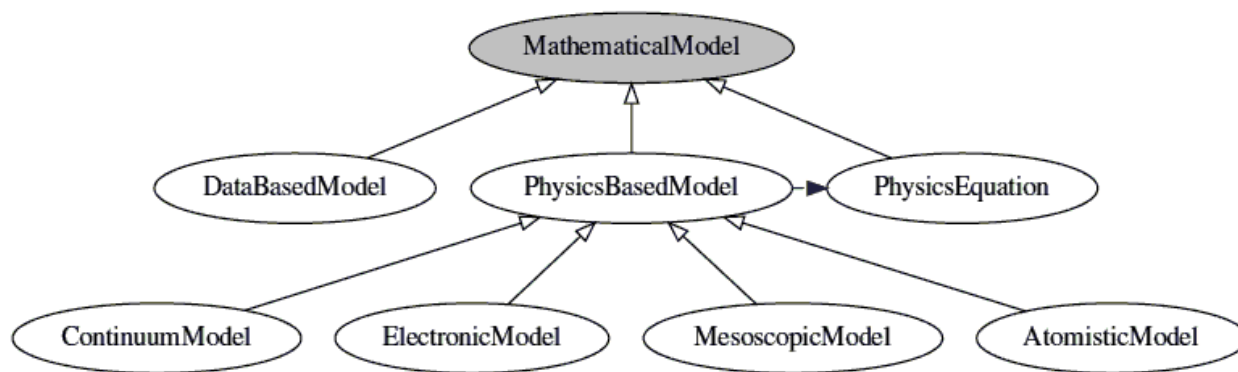


Рис. 6. Иерархия классов, отражающая виды математических моделей (из описания онтологии ЕММО [27])

Подробное описание всех классов с истолкованием их смысла и применения можно найти в документации по онтологии [27]. На разных уровнях промежуточного слоя, берущего начало с класса *Perspective* (рис. 4), представлены всевозможные элементы, необходимые в ходе моделирования: физические свойства, процессы, математика и программы, метрология, геометрические фигуры и проч.

Наличие готовой структуры в виде ядра и промежуточного уровня резко облегчает главную задачу ЕММО – обеспечить интеграцию данных и моделей при охвате любых, достаточно узких областей знания. Для ее решения структура ЕММО настраивается на соединение с прикладной онтологией узкого профиля, чьи классы включаются в ЕММО на правах нижнего уровня. Методика подобной процедуры подробно освещается в документе [28] на простом примере моделирования прочностных свойств сварного элемента. В работе [29] описана аналогичная процедура включения в онтологию концепций, отражающих процессы механического тестирования.

В репозитории [26], где хранятся различные версии ЕММО, среди прочих приведено несколько версий узкого профиля, например, включающих элементы кристаллографии или данные по микроструктуре вещества. На том же сайте находятся программные средства в форме пакета на языке Python [30], который использует разработанный в статье [31] метод онтологического программирования. На его основе широкий круг пользователей получает возможность практически освоить разработку онтологий для множества задач материаловедения и физики.

ЗАКЛЮЧЕНИЕ

В настоящей статье рассмотрена принципиально новая организация исследований, обусловленная их ориентацией на использование обширных массивов данных и методов их анализа и обработки. Широта решаемых на этой основе проблем в науке о материалах позволяет сделать выводы, справедливые вне зависимости от конкретной области.

Организирующим началом в постановке исследований служит то, что принято называть «инфраструктурой данных», т. е. унифицированная и интегриро-

ванная система из программных средств хранения данных и моделирования, а также многочисленных сервисов по аренде ресурсов, тренингу специалистов, распространению стандартов и проч. Эта инфраструктура создается на основе новой бизнес-модели, так называемой двусторонней платформе, соединяющей на едином плацдарме заказчиков из промышленности и экспертов в предметной области. При этом все участники рынка получают многочисленные выгоды и преимущества за счет обширного портфеля заказов, доступности средств их реализации и штата специалистов, способных адаптировать модели и программные средства для анализа конкретных материалов и технологий, применяемых в индустрии.

Предоставление услуг в сфере высоких технологий (математическое моделирование, обработка больших данных, облачный сервис) потребовали выстроить все управление и координацию отдельных звеньев инфраструктуры на основе системы онтологий, обеспечивающих стандартизацию понятийного аппарата и логической структуры используемых данных. Для управления рынком VIMMP потребовалась не одна, а экосистема узко-ориентированных онтологий, характеризующих сервисы, модели и взаимодействие пользователей. Причем собственно физическим моделям и последовательности действий при моделировании посвящены онтологии ЕММО и OSMO, а другие ориентированы на частные аспекты в жизни виртуального рынка, например, выбор модели для заказов от промышленности (ММО), учебные сервисы (OTRAS) или коммуникации между участниками рынка (VICO).

Созданию онтологий всегда предшествует разработка или использование действующих стандартов, которые регламентируют терминологию и классификаторы. Применительно к материаловедению, стандарт *de facto* (не являющийся нормативным документом), предложенный в документе [18] включает классификационную схему физических моделей, а также стандарт MODA, предъявляющий требования к описанию проектов. Уникальность обоих стандартов в их исключительной широте, позволяющей охватить все виды моделей, начиная с методов квантовой химии и вплоть до феноменологии, применяемой при расчете прочности или теплообмена. Адекват-

ватность этих стандартов подтверждена отчетностью по большой группе проектов, представленной в том же документе.

Опыт работы виртуального рынка в науке о материалах показал, что бизнес-модель в виде двусторонней платформы, система онтологий и установленные стандарты формируют ядро, способное обеспечить эффективность исследований и разработок в любой из дисциплин, связанных с интенсивным использованием данных.

СПИСОК ЛИТЕРАТУРЫ

1. Когаловский М.Р. Энциклопедия технологий баз данных. – Москва: Финансы и статистика, 2002. – 800 с.
2. Срничек Н. Капитализм платформ / пер. с англ. и науч. ред. М. Добряковой. 2е изд. – Москва: Изд. дом Высшей школы экономики, 2020. – 128 с.
3. Valkovari Katri. VTT. Novel business models for digital platforms // EMMC International Workshop 2019. Feb 25-27, 2019. – Vienna/Austria. – URL: <https://emmc.info/emmc-international-workshop-2019-workshop-presentations/>
4. Скворцов Н.А., Калиниченко Л.А., Ковалев Д.Ю. Концептуальное моделирование предметных областей с интенсивным использованием данных // Труды XVIII Международной конференции DAMDID/RCDL'2016 «Аналитика и управление данными в областях с интенсивным использованием данных» (Ершово, 11-14 октября 2016 г.). – URL: <http://ceur-ws.org/Vol-1752/paper03.pdf>
5. Кузнецов О.П., Суховеров В.С., Шипилина Л.Б. Онтология как систематизация научных знаний: структура, семантика, задачи // Труды конференции «Технические и программные средства систем управления, контроля и измерения». – Москва, 2010. – С. 000762-000773.
6. Еркимбаев А.О., Зицерман В.Ю., Кобзев Г.А. Интенсивное использование цифровых данных в современном естествознании // Научно-техническая информация. Сер. 2. – 2017. – № 9. – С. 9-22; Erkimbaev A.O., Zitserman V.Yu., Kobzev G.A. The Intensive Use of Digital Data in Modern Natural Science // Automatic Documentation and Mathematical Linguistics. – 2017. – Vol. 51, № 5. – P. 201-213.
7. EMMC-CSA – GA N°723867. Report on Sessions on Interoperability and integration within the EMMC International Workshop 2019, FEB 25-27, 2019, Vienna Austria. – URL: https://emmc.info/wp-content/uploads/2019/06/EMMC-CSA_D2.12_M31_vfinal-PU-WEB1706.pdf
8. Hagi A., Wright J. Multi-sided platforms // International Journal of Industrial Organization. – 2015. – Vol. 43. – P. 162-174.
9. Коваленко А.И. Проблематика исследований многосторонних платформ // Современная конкуренция. – 2016. – Т.10, № 3(57). – С. 64-90.
10. Панышин Б. Цифровая экономика: особенности и тенденции развития // Наука и инновации. – 2016. – Т. 3., № 157. – С. 17-20.
11. Himanen L., Geurts A., Foster A.S., Rinke P. Data-Driven Materials Science: Status, Challenges, and Perspectives // Advanced Science. – 2019. – Vol. 6. – 1900808.
12. Materials Modelling is Key to Innovation. – URL: <https://emmc.info/the-emmc-mission/>
13. Virtual Materials Marketplace. – URL: <http://vimmp.eu>
14. Horsch M.T., Chiacchiera S., Seaton M.A., et al. Introduction to the VIMMP ontologies. Technical report. – Brussels, Belgium: EMMC ASBL, 2020. – URL: <https://zenodo.org/record/3936796#.YG7pfTgzaUk>
15. Horsch M.T., Chiacchiera S., Seaton M.A., et al. Ontologies for the Virtual Materials Marketplace // Künstl Intell. – 2020. – P. 1-6. – URL: <https://doi.org/10.1007/s13218-020-00648-9>
16. Translators Guide. – URL: <https://emmc.info/wp-content/uploads/2018/01/Translators-Guide.pdf>
17. CEN Workshop Agreement. European Committee for Standardization. Materials modelling – terminology, classification and metadata. (2017). CWA 17284. – URL: <https://www.cen.eu/news/workshops/Pages/WS-2017-012.aspx>
18. What makes a material function? Let me compute the ways Modelling in H2020 LEIT-NMBP. Programme Materials and Nanotechnology projects Sixth version / ed. Anne de Baas, 2017. Directorate-General for Research and Innovation, Industrial Technologies. – URL: <https://emmc.info/romm6/>
19. ISO/IEC 19510:2013. Information technology — Object Management Group Business Process Model and Notation. – URL: <https://www.iso.org/standard/62652.html>
20. de Baas Anne. Materials Modelling Data and Documentation: terminology, classification and ontology towards Digital Single Market // Interoperability in Materials Modelling – IntOP2017/ – URL: <https://emmc.info/events/intop17/>
21. Абгарян К.К. Многомасштабное моделирование в задачах структурного материаловедения. – Москва: МАКС Пресс, 2017. – 284 с.
22. Новая философская энциклопедия. Статья «Семiotика» / научно-ред. совет: В.С. Стёпин, А.А. Гусейнов, Г.Ю. Семигин, А.П. Огурцов. 2-е изд., испр. и допол. – Москва: Мысль, 2016. – Т. 1-4. – URL: <https://biblioclub.ru/index.php?page=dict&termin=175995>
23. Achille Varzi. Mereology. Stanford Encyclopedia of Philosophy. – URL: <https://plato.stanford.edu/entries/mereology/>
24. Smith Barry. Classifying Processes: An Essay in Applied Ontology // Ratio (Oxf). – 2012. – Vol. 25, № 4. – P. 463-488.
25. Simperler A. Introduction to EMMO. Part II. Online 29.4.2019. – URL: https://emmc.info/wp-content/uploads/2019/04/Part_2_EMMO_Intro.pdf
26. Elementary Multiperspective Material Ontology (EMMO). – URL: <https://github.com/emmo-repo/EMMO#readme>

27. Версии онтологии EMMO и документация к ним. – URL: <https://emmo-repo.github.io/>
28. EMMC Documentation on “Design and Implementation of metadata schema for syntactic and semantic interoperability”, Published on 2019. – URL: https://wp-content/uploads/2019/08/EMMC-CSA_-_D2.9_M30_vfinal-PU_WEB.pdf
29. Morgado J.F., Ghedini E., Goldbeck G., et al. Mechanical testing ontology for digital-twins: A roadmap based on EMMO // Proceedings of the International Workshop on Semantic Digital Twins (SeDiT 2020). – URL: <http://eur-ws.org/Vol-2615/>
30. EMMO – Python API for the European Materials & Modelling Ontology. – URL: <https://github.com/emmo-repo/EMMO-python>
31. Lamy Jean-Baptiste. Owlready: Ontology-oriented programming in Python with automatic classification and high level constructs for biomedical ontologies // Artificial Intelligence In Medicine. – 2017. – Vol. 80. – P.11-28.

Материал поступил в редакцию 06.10.2021

Сведения об авторах

ЕРКИМБАЕВ Адильбек Омирбекович – кандидат технических наук, зам. зав. лабораторией теплофизических баз данных Объединенного института высоких температур Российской академии наук (ОИВТ РАН), Москва
e-mail: adilbek@ihed.ras.ru

ЗИЦЕРМАН Владимир Юрьевич – кандидат физико-математических наук, ведущий научный сотрудник лаборатории теплофизических баз данных ОИВТ РАН
e-mail: vz1941@mail.ru

КОБЗЕВ Георгий Анатольевич – доктор физико-математических наук, главный научный сотрудник лаборатории теплофизических баз данных ОИВТ РАН
e-mail: gkbz@mail.ru

КОСИНОВ Андрей Владимирович – инженер-программист опытно-конструкторского бюро ОИВТ РАН
e-mail: kosinov@gmail.com