

# Коммюнике Онтологического Саммита 2016<sup>1</sup>

## ОНТОЛОГИИ В СЕМАНТИЧЕСКИ ИНТЕРОПЕРАБЕЛЬНЫХ ЭКОСИСТЕМАХ<sup>2</sup>

Главные редакторы: **Donna Fritzsche, Michael Gruninger**

Редакторы: **Ken Baclawski, Mike Bennett, Gary Berg-Cross, Leo Obrst, Todd Schneider, Ram Sriram, Mark Underwood, Andrea Westerinen**

Introduction

Semantic Interoperability Overview

What is Semantic Interoperability?

The Role of Ontologies

Semantic Heterogeneity

Kinds of Ontologies in the SI Ecosystem

Overview

Upper Ontologies

Reference Ontologies

Application and Local Ontologies

Bridge Ontologies

Common Metadata Templates and Metadata Schema

Architectural Developments

Semantic Data Lakes

Information Modeling for Federation

The Cloud Presents New Opportunities

Methods which Facilitate Semantic Interoperability

Ontology Reuse and Modular Design

Vocabulary Harmonization

Ontology Mapping

Ontology Design Patterns

Toward the Right Mix

Engineering and Manufacturing

Healthcare and Bioinformatics

Earth Sciences

Finance and Retail

Designing for Semantic Interoperability

Which Tools and Environments are Necessary?

Ontologies and Communities

Lessons Learned

A Roadmap for the Future

A Vision for Semantically-Aware Ecosystems

Communication Strategy

Testbeds for Semantic Interoperability

Education Workshops

Funding and Ongoing Support

In Closing

References

---

<sup>1</sup> Сокращенный перевод *Ontology Summit 2016 Communique. Ontologies within Semantic Interoperability Ecosystems.*  
<http://ontologyforum.org/index.php/OntologySummit2016/Communique>

<sup>2</sup>Новая, разрабатываемая международным сообществом отрасль знаний «онтологический инжиниринг» определяет для всех свой новый словарь терминов. Как в своё время термины кибернетика и компьютер «с трудом» входили в русскоязычный оборот, так и сейчас обилие «труднопереводимых» терминов интероперабельность, экосистемы и др. фактически вторгаются в русскоязычный оборот, расширяя его в силу отсутствия точного русскоязычного синонима. Также обстоит дело и с компьютерным слэнгом, в который уже вошли «облачные вычисления» и вот в этом документе - «семантические озёра данных». *Прим.ред.*

## Введение

Способность к взаимодействию между системами – привычное требование в современной цифровой информационной среде.

Предприятия, организации и исследовательские группы стремятся предоставлять услуги наилучшего качества, сводить к минимуму информационные издержки, сокращать расходы и создавать инновационные проекты, используя:

- Интернет вещей;
- умные сети электроснабжения;
- интеллектуальные агенты и машинное обучение;
- персонализированные услуги и адресную доставку контента;
- методы интеллектуального анализа данных.

Требуется как синтаксическая, так и семантическая интеграция между системами и приложениями. Однако, на практике семантическая интероперабельность (СИ) труднодостижима.

Онтологии и связанные с ними системы интеллектуального вывода (оба термина использованы в самом широком смысле слова) являются ключевыми элементами в обеспечении семантической интеграции и интероперабельности. Однако возникает несколько вопросов. Как мы определяем инструменты, методологии и программные платформы, поддерживающие эту интероперабельность? Что требуется для достижения оптимально продуктивного взаимодействия систем и приложений. Как нам строить диалог при обсуждении роли онтологий в обеспечении интероперабельности в гетерогенной среде?

В рамках Онтологического саммита 2016 мы изучали то, как онтологии и онтологические методы могут помочь в обеспечении СИ. В этом документе представлен обзор проблем, возникающих при реализации СИ. Мы обсудим онтологические инструменты и методологии обеспечения СИ и представим конспективное изложение свежих работ по этой теме, затрагивающих четыре обширных предметных области (PrO): здравоохранение/фармацевтика, земледелие, машиностроение и финансы/ритейл. В завершение коммюнике мы приведем подходы к проектированию, стратегии и направления для будущих исследований в этой отрасли.

## Обзор семантической интероперабельности

### Что такое семантическая интероперабельность?

*Интероперабельность* - это способность двух или более сущностей взаимодействовать для того, чтобы достичь одной или нескольких целей или решить какие-либо тактические задачи, в то время как интеграция - это сближение систем, вплоть до объединения, опять же для решения тактических задач (например, повышение функциональности, скорости, снижение издержек, управление рисками и т.д.). В контексте интероперабельности, семантика - это базовая интерпретация в некотором контексте. Семантическая интероперабельность - это способность получателя информации интерпретировать или понимать содержание примерно в том смысле, в котором его посылал отправитель для достижения общих целей (например, контекст для и от информации).

### Роль онтологий

В качестве ключевой технологии поддержки СИ предлагается использование правильно организованных и развитых онтологий. Зарождение прикладной онтологии в проекте «Shareable and Reusable Knowledge Base» демонстрирует пути решения проблемы СИ. В этой программной среде онтологии предоставляют набор терминов и отношений между ними, вместе с машиночитаемой спецификацией, и толкования (значения) этих терминов. Их задача – поддержка СИ информационных систем и источников данных путем использования терминологии, содержащейся в их онтологиях. СИ и интеграция могут быть достигнуты за счет:

- установления базовых семантических представлений через онтологии (уровень классов) и их баз знаний (уровень сущностей);
- определения семантических соответствий и трансформаций между онтологиями;
- использования алгоритмических методов, которые определяли бы семантическое сходство для установления связей между онтологиями.

Совместное использование онтологий и программного обеспечения для поиска семантических соответствий снижает потери семантики (смысла) в обмене информацией между гетерогенными приложениями.

Онтологический Саммит 2009 года предложил Среду использования онтологий, которая включала несколько тем, относящихся к СИ. Одна называлась «интеграция информации», когда множественные информационные ресурсы объединяются с использованием онтологий для поиска концептов со сходными значениями. Примерами этого подхода могут служить сбор данных и объединение данных. Вторым способом использования онтологий является обеспечение программной интероперабельности, когда программные системы взаимодействуют путем обмена сообщениями, которые составлены с использованием онтологий. Каждая система использует онтологию (либо свою собственную, либо пользуется набором общих онтологий) для перевода полученных сообщений.

### Семантическая гетерогенность

Проблема семантической гетерогенности нередко встречается в современном машиностроении и производстве, что делает её немаловажной при обсуждении проблем СИ. Многогранное понятие, она возникает из-за необходимости обмениваться контентом и данными внутри и между несколькими информационными экосистемами. Свидетельства этому мы обнаружили во всех ПрО, рассмотренных на Онтологическом Саммите 2016.

В информационных экосистемах одновременно используются устройства, датчики, программные продукты и источники данных. В разных системах информация о ПрО:

- описана в множественных схемах;
- описана с использованием словарей с разной и локально развитой семантикой;
- реализована в разных языках разметки;
- основана на моделях, созданных на базе разных концепций.

Данные также представлены с разными уровнями детализации, чья семантика основывается на разных моделях. В результате, зачастую информация не может быть передана между программами, и источники данных не могут быть продуктивно объединены. Дополнительно, может страдать взаимодействие конечных пользователей, инженеров по знаниям, экспертов ПрО.

Продукты становятся все более сложными и на разных этапах своего жизненного цикла существуют в разных информационных и физических средах: concept, design, development, testing, manufacturing/deployment, operation, upgrade, decommissioning. Важно понимать сущности, вовлеченные в процесс, и разнообразные физические и программные среды в которых они существуют или которые проходят в своём жизненном цикле (такие как CAD, CAM, испытательный стенд или фабрика), и, что самое важное, соответствующие отношения. В каждой среде, или на каждом этапе жизненного цикла, принимаются решения, затрагивающие последующие (и иногда предыдущие) этапы или среды.

Проектирование и производство таких сложных объектов требует совместного участия и обмена данными, необходимыми для принятия решений, от нескольких отраслей. Однако, в технических отраслях сформировались собственные концепты, отношения, терминология, парадигмы и инструменты для решения собственных задач. В результате этого процесса строгая семантика и специфические трактовки стали широко употребимы среди профессионалов в этой области и вошли во множество стандартов, спецификаций и инструментов (как программных, так и аппаратных, например, станочное оборудование). Таким образом, при возникновении необходимости совместного труда и обмена информацией между несколькими дисциплинами интероперабельность терминологии, данных и инструментов превращается в проблему. Такие проблемы приводят к потере времени, возникновению ошибок и увеличению стоимости разработки, производства или эксплуатации.

Отсутствие интероперабельности дорого обходится, так как оно повышает стоимость внедрения и ухудшает пользовательский опыт. Согласно оценке Office of Financial Research отсутствие общего языка в финансовой отрасли вызывает многомиллиардные убытки. Согласно отчету Center for Medicare and Medicaid Services, в 2014 году США потратили на здравоохранение около 3 триллионов долларов и к концу декады эта сумма может удвоиться. Предполагается, что путем эффективного использования информационных технологий в здравоохранении удастся значительно сохранить операционные расходы и спасти дополнительно тысячи жизней в год. Интероперабельные системы способны значительно повысить качество взаимодействия врача и пациента.

Рассмотрим иерархию интеграции, необходимой для достижения интероперабельности. На нижнем конце спектра находится синтаксическая интероперабельность, в середине – структурная и на вершине – семантическая. Как правило, большинство представителей промышленности и практиков ПрО сходятся на использовании нижнего и среднего уровней интеграции с общими протоколами и форматами данных для обеспечения обмена данными. Существует мнение, что СИ достижима путём стандартизации, например, через

управляемые словари. Однако, надёжная семантическая интеграция требует одинаковой интерпретации сообщений и пересылаемых данных, то есть их смысл остаётся неизменным в процессе обмена между разными ПрО и множественными системами в рамках объединенных экосистем.

## **Виды онтологий в семантически интероперабельных экосистемах**

### **Обзор**

В рамках Онтологического Саммита 2007 было исследовано множество семантического контента (включая таксономии, тезаурусы, тематические карты, концептуальные модели и формальные онтологии, описанные на различных логических языках) которое составляет артефакты, обычно называемые онтологиями. Несмотря на то, что онтологии способны решить проблему СИ, разнообразие онтологий и семантического контента, описанного на разных логических языках, создаёт возможность только усугубить проблему при неконтролируемом использовании. Несмотря на рост количества и качества онтологий, в настоящее время ситуация всё ещё может быть описана выражением «семантический беспорядок». Ситуация сохраняется из-за того, что информация о ПрО гетерогенна и описывается в форме:

- разнообразных схем;
- разных словарей и языков разметки;
- онтологий с различными уровнями детализации данных и с использованием различных концептуализаций.

Есть довольно много онтологий, распределённых по спектру семантической формализации, а также по степени детальности и полноты. Как использовать этот семантический контент всего спектра онтологий?

Примером, иллюстрирующим применение и взаимное обогащение онтологий разного уровня детализации, может служить онтология NASA's Semantic Web for Earth and Environmental Terminology (SWEET), которая содержит около 6000 концептов, распределённых по более чем 200 отдельным модульным онтологиям. Такие онтологии наращивают усилия сообщества по разработке стандартных словарей в ПрО для поддержки интероперабельности данных и систем. SWEET помогает в решении таких задач как семантическая разметка, однако она содержит мало аксиом для поддержки логического вывода и её необходимо дополнять каждый раз, когда она используется для решения продвинутых задач. В целом признаётся, что некоторые ранние онтологии бедны с точки зрения формальной семантики и что более аксиоматизированные онтологии, отражающие структуру знаний о ПрО, могут помочь в обходе этого ограничения. С другой стороны, многие термины из биомедицинской или естественнонаучной отраслей иногда «не совсем научны». Примерами могут служить идеи о реке, канале, водоёме или связи симптомов и болезни. Учёные или медики просто не имеют полного и ясного понимания этих понятий, особенно в связи с другими концептами, которые и составляют реальность в виде системы. Это отчасти отражает сложность реальности и состояние текущего ограниченного понимания науки.

### **Онтологии верхнего уровня**

Проведение первого Онтологического Саммита 2006 года было продиктовано необходимостью решения семантических различий между онтологиями, в частности, в отношении их онтологических обязательств.

Онтологии верхнего уровня, такие как Basic Formal Ontology (BFO), Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE), UpperCYC, ISO 15926, COSMO, Suggested Upper Merged Ontology (SUMO) - это инструменты, которые могли бы помочь в решении проблемы семантической гетерогенности, так как они однозначно аксиоматизируют концепты, используемые в большом количестве специализированных онтологий. Полное или частичное использование концептов верхней онтологии и её методов моделирования в низших онтологиях может помочь свести к минимуму введение дополнительных смысловых несоответствий в экосистеме.

Использование онтологий верхнего уровня может помочь смягчить проблему СИ, но остаются несколько серьёзных проблем.

- Существует несколько конкурирующих онтологий верхнего уровня.
- Многие онтологии верхнего уровня сложны для понимания, содержат сложные аксиомы и абстракции, слишком далеко оторванные от реальных данных.
- Они должны быть согласованы с предметной областью, имеющимися ссылками, приложениями и частными онтологиями.
- Онтология верхнего уровня, особенно построенная «сверху вниз», может налагать онтологические ограничения, неприемлемые для всех заинтересованных сторон, имеющих «местные» словари и значения.

- Как и системы искусственного интеллекта в прошлом, онтологии верхнего уровня могут быть слишком хрупкими, то есть внесение небольших изменений может вызывать затруднения или ставить под угрозу семантику.

### Эталонные (или ссылочные) онтологии

Эталонные онтологии отражают знания базового уровня о широкой ПрО или семантический консенсус индустриальной отрасли. Они происходят из попыток отразить глубокие знания фундаментальной науки в виде принципов, предназначены для повторного использования и жестко не привязаны к конкретным случаям использования в приложениях и налагаемым ими требованиям. По замыслу, они созданы для облегчения интеграции между системами, хранилищами и источниками данных. Вместо того, чтобы служить в качестве верхней онтологии, которая выступает посредником между другими онтологиями, эталонная онтология служит средством для связи терминологий нескольких информационных систем и данных в наборе общих понятий. Классическим примером такой онтологии является The Foundational Model of Anatomy (FMA), состоящая из примерно 75 000 классов, 120 000 терминов и 168 типов отношений. FMA - это эталонная онтология, представляющая структуру человеческого тела. В случае правильной реализации, коллекция эталонных онтологий может рассматриваться как ортогональные (непересекающиеся) интероперабельные ресурсы. Это примерно соответствует модели, принятой в Open Biological and Biomedical Ontologies (OBO) Foundry, которая состоит из онтологий с молекулярного уровня до уровня организма, описывая как биологическую структуру, так и функциональную.

### Онтологии приложений и локальные онтологии

Онтология приложения - это «онтология, разработанная для конкретной задачи или приложения и сфера которой определяется посредством тестируемых сценариев использования» (Malone & Parkinson 2010). Ключевым моментом является то, что они созданы с для удовлетворения требований проекта. Они могут быть применимы к локальной предметной области или охватывать несколько смежных областей. В том случае, если локальные онтологии создаются без связи с более общими онтологиями, их довольно непросто интегрировать с другими онтологиями. В результате, некоторый уровень интеграции достигим установкой соответствия между концептами онтологий различных приложений. В свою очередь, систематизируя такое установления соответствия, возможно создать основу для эталонной онтологии.

### Онтологии связи

Родственные эталонным онтологиям (которые связывают терминологии нескольких систем), онтологии связи, как правило, используются в качестве посредника между конкретными понятиями множественных онтологий. Посредничество может быть простым и достигаться путём сопоставления между подобными понятиями. Часто, однако, связь концептов требует добавления новых концептов, которые могут отсутствовать в рассматриваемых онтологиях. Добавленный концепт обеспечивает соединение двух или более связанных с ним понятий между онтологиями. Примером такого концепта может служить процесс, в котором участвуют оба концепта в объединяемых онтологиях и который при этом мог быть не описан ни в одной из них. Умный брокер, например, может знать, что понятие «проводимость» связывает «солёность» и «морскую воду». Когда интеллектуальный брокер получает запрос на данные о солёности, он будет отображать данные о проводимости морской воды, если пользователь хочет видеть связанные термины в результатах запроса.

Несмотря на тот факт, что задача связи между онтологиями обычно решается онтологиями верхнего уровня, существуют случаи, когда онтологии связи позволяют обнаруживать общности между различными локальными онтологиями в одной ПрО. Существующие стандарты, онтологии, связанные словари и модели (например, GeoScience), как правило разрабатывались изолированно, что порождает значительные проблемы при попытке их объединения. Онтологии выступают в качестве клея, помогающего их соединять и гармонизировать. Онтология связи может быть уже эталонной онтологией и содержать меньшее количество аксиом. Использование онтологии связи часто делает очевидной необходимость использования более полной модели, например, в виде эталонной онтологии или онтологии приложения с более широкими требованиями.

### Общие шаблоны метаданных и схемы метаданных

## **Архитектурные Разработки**

- Озёра семантических данных
- Информационное моделирование для сотрудничества
- Облако открывает новые возможности

## **Методы, которые облегчают семантическую интероперабельность**

- Повторное использование онтологий и модульный принцип создания
- Гармонизация словаря
- Управление онтологиями
- Модели проектирования онтологий

## **К правильной комбинации**

- Проектирование и производство
- Здравоохранение и биоинформатика
- Землеведение
- Финансы и розничная торговля

## **Проектирование для достижения семантической интероперабельности**

### Какие инструменты и среды необходимы?

Участники Саммита сошлись во мнении о необходимости лучших программных инструментов поддержки интеграции концептов или данных. Почти для всех методов, рассмотренных в предыдущем разделе, отсутствуют адекватные среды разработки и технического обслуживания. Таким образом, существует необходимость в новых инструментах и постоянном совершенствовании существующих, которые обеспечивают разработку и поддержку озёр семантических данных, федеративных систем и шаблонов проектирования онтологий. Точно так же есть постоянная потребность в инструментах и методах, которые облегчат: повторное использование онтологий и модульный принцип проектирования, гармонизации лексики и установления соответствия концептам онтологии. В следующих разделах мы приводим подробное описание дополнительной функциональности, которая поможет в процессе проектирования.

- Онтология и Сообщества
- Извлеченные уроки

## **План на будущее**

- Представление о семантически ответственных экосистемах
- Стратегия общения
- Инструменты оценки семантической интероперабельности
- Образовательные семинары
- Финансирование и постоянная поддержка

## **В завершение**

Решение задачи СИ представляет серьезный вызов в широком диапазоне предметных областей. В ходе Онтологического саммита 2016 были представлены решения, которые описывают как онтологии (во всем спектре онтологий) используются для решения проблемы семантической гетерогенности. В частности,

разрабатываются специальные онтологии для поддержки обмена данными между множественными системами. В то же время, разрабатываются или углубляются дополнительные методы, которые облегчают решение задачи обеспечения СИ. Эти методы варьируются от отображения онтологий и гармонизации лексики до разработки модульных повторно используемых онтологий и шаблонов проектирования онтологий. В области архитектуры развитие включают в себя озёра семантических данных. Несмотря на достигнутый прогресс, проблемы всё ещё остаются, достичь интероперабельности всё ещё тяжело. На Онтологическом саммите 2016 был достигнут широкий консенсус относительно важности усовершенствования программных средств и сред, необходимых для поддержки интеграции концептов и данных. Мы обсудили множество социально-технических проблем, затрудняющих использование онтологий в поддержке СИ. Для решения этих проблем были сформированы несколько предложений в план на будущее, касающихся, в том числе, создания тестовых стендов семантической совместимости, активизации усилий по взаимодействию, проведению образовательных семинаров, а также постоянную поддержку скоординированного развития онтологий и отображений онтологий.

## Источники

[https://financialresearch.gov/frac/files/FRAC\\_DTS\\_Instrument\\_Database\\_Recommendation.pdf](https://financialresearch.gov/frac/files/FRAC_DTS_Instrument_Database_Recommendation.pdf)  
<http://www.cms.gov>  
<https://en.wikipedia.org/wiki/Semantics>  
<https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html>  
<https://sweet.jpl.nasa.gov/>  
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839690/>  
<http://si.washington.edu/projects/fma>  
<http://www.edmcouncil.org/financialbusiness>  
<http://www.omg.org/hot-topics/finance.htm>  
<http://ontogenesis.knowledgeblog.org/295>  
<http://schema.org/docs/about.html>  
<http://metadatascenter.org/>  
[http://www.omgwiki.org/architecture-ecosystem/doku.php?id=semantic\\_information\\_modeling\\_for\\_federation\\_rfp](http://www.omgwiki.org/architecture-ecosystem/doku.php?id=semantic_information_modeling_for_federation_rfp)  
[https://s3.amazonaws.com/ontologforum/OntologySummit2016/2016-03-10\\_Healthcare/The-Semantic-Data-Lake-in-Healthscience--ParsaMirhaji.pdf](https://s3.amazonaws.com/ontologforum/OntologySummit2016/2016-03-10_Healthcare/The-Semantic-Data-Lake-in-Healthscience--ParsaMirhaji.pdf)  
<http://www.commonwl.org/draft-3/SchemaSalad.html>  
[http://www.omgwiki.org/architecture-ecosystem/lib/exec/fetch.php?media=semantic\\_information\\_modeling\\_for\\_federation\\_overview.pdf](http://www.omgwiki.org/architecture-ecosystem/lib/exec/fetch.php?media=semantic_information_modeling_for_federation_overview.pdf)  
<http://ontolog.cim3.net/OntologySummit/2014/communicate.html>  
[http://wiki.goodrelations-vocabulary.org/GoodRelations\\_and\\_schema.org](http://wiki.goodrelations-vocabulary.org/GoodRelations_and_schema.org)  
<http://www.edmcouncil.org/semanticsrepository/index.html>  
<http://linkeddata.org/>  
<http://www.opengeospatial.org/standards/om>  
<https://www.w3.org/2005/Incubator/ssn/ssnx/ssn>  
<http://www.omg.org/spec/DOL/1.0/Beta1/index.htm>  
[http://ceur-ws.org/Vol-1081/womo2013\\_invited\\_paper\\_1.pdf](http://ceur-ws.org/Vol-1081/womo2013_invited_paper_1.pdf)  
<https://global.oup.com/academic/product/a-pattern-language-9780195019193?cc=us&lang=en&>  
<https://nrf.com/>  
<http://www.searchmetrics.com/news-and-events/schema-org-in-google-search-results/>  
<http://www.obofoundry.org/>  
<http://www.allotrope.org/#!organization/c20k4>  
[https://ec.europa.eu/esco/portal/escopedia/European\\_Skills%252C\\_Competerences%252C\\_Qualifications\\_and\\_Occupations\\_%2528E\\_SCO%2529](https://ec.europa.eu/esco/portal/escopedia/European_Skills%252C_Competerences%252C_Qualifications_and_Occupations_%2528E_SCO%2529)  
<http://vivoweb.org/info/about-vivo>  
[http://ontologforum.org/index.php/ConferenceCall\\_2016\\_04\\_21](http://ontologforum.org/index.php/ConferenceCall_2016_04_21)  
[http://ontologforum.org/index.php/ConferenceCall\\_2016\\_02\\_18](http://ontologforum.org/index.php/ConferenceCall_2016_02_18)  
<https://www.ncoic.org/technology/technical-products/scope-workshops>  
 Arp, Robert; Smith, Barry; Spear, Andrew D. (2015-08-28), Building Ontologies with Basic Formal Ontology, The MIT Press, Kindle Edition.