
Концептуальная стадия информационной интеграции в архитектуре межведомственных услуг

Зиндер Евгений Захарович

Фонд ФОСТАС, президент

e-mail: ezinder@fostas.ru

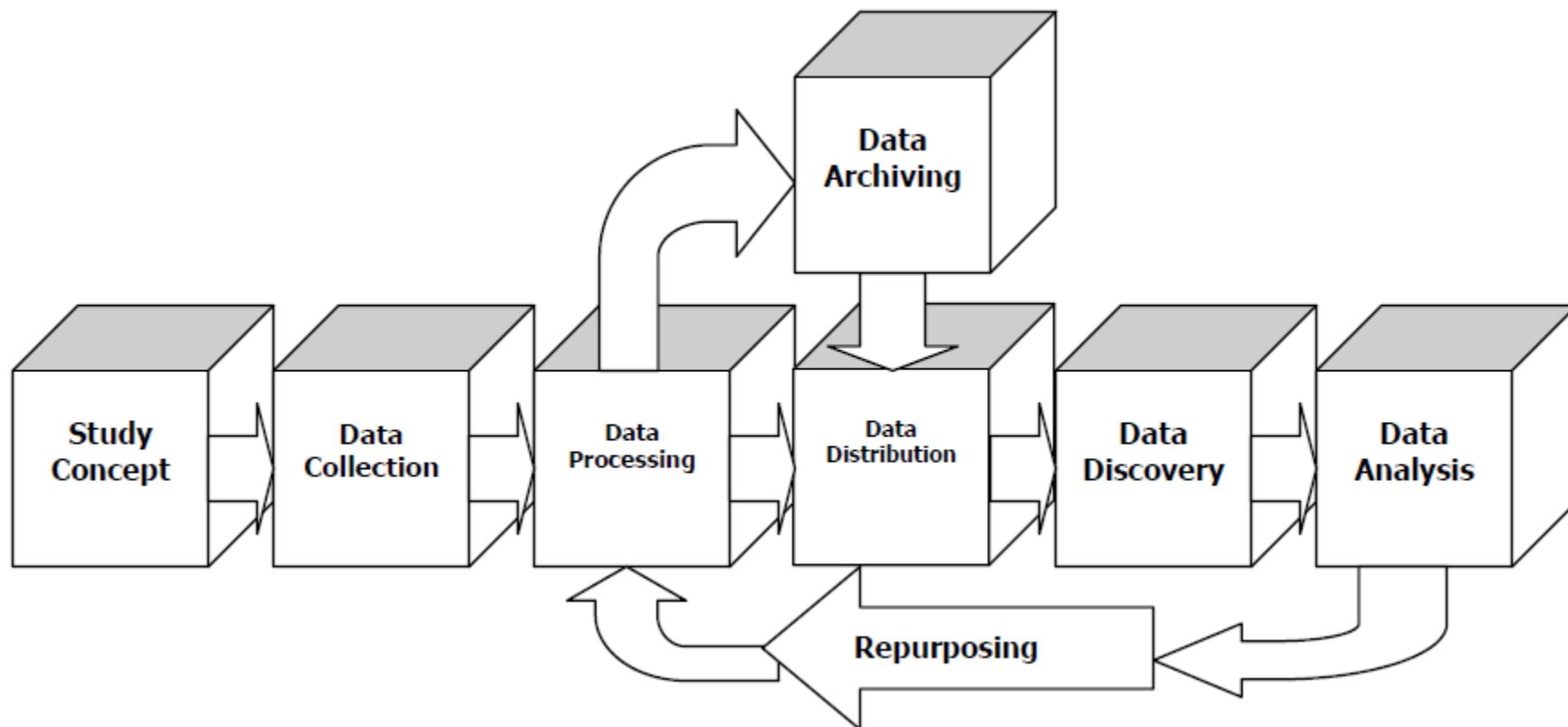
На этом семинаре 20 октября рассматривался цикл информационных работ для интеграции систем –

- **по DDI**
- **по OASIS-CAM, по NIEM**
- **обобщенный**

Были показаны

- **основные источники неоднозначных и ошибочных информационных решений**
- **их места в самой начальной части работ, еще не связанных с ИКТ, но особенно тяжелых по последствиям**
- **гибридные способы и оперативного разрешения в ходе эксплуатации**

Одна из моделей цикла информационных работ - Data Documentation Initiative



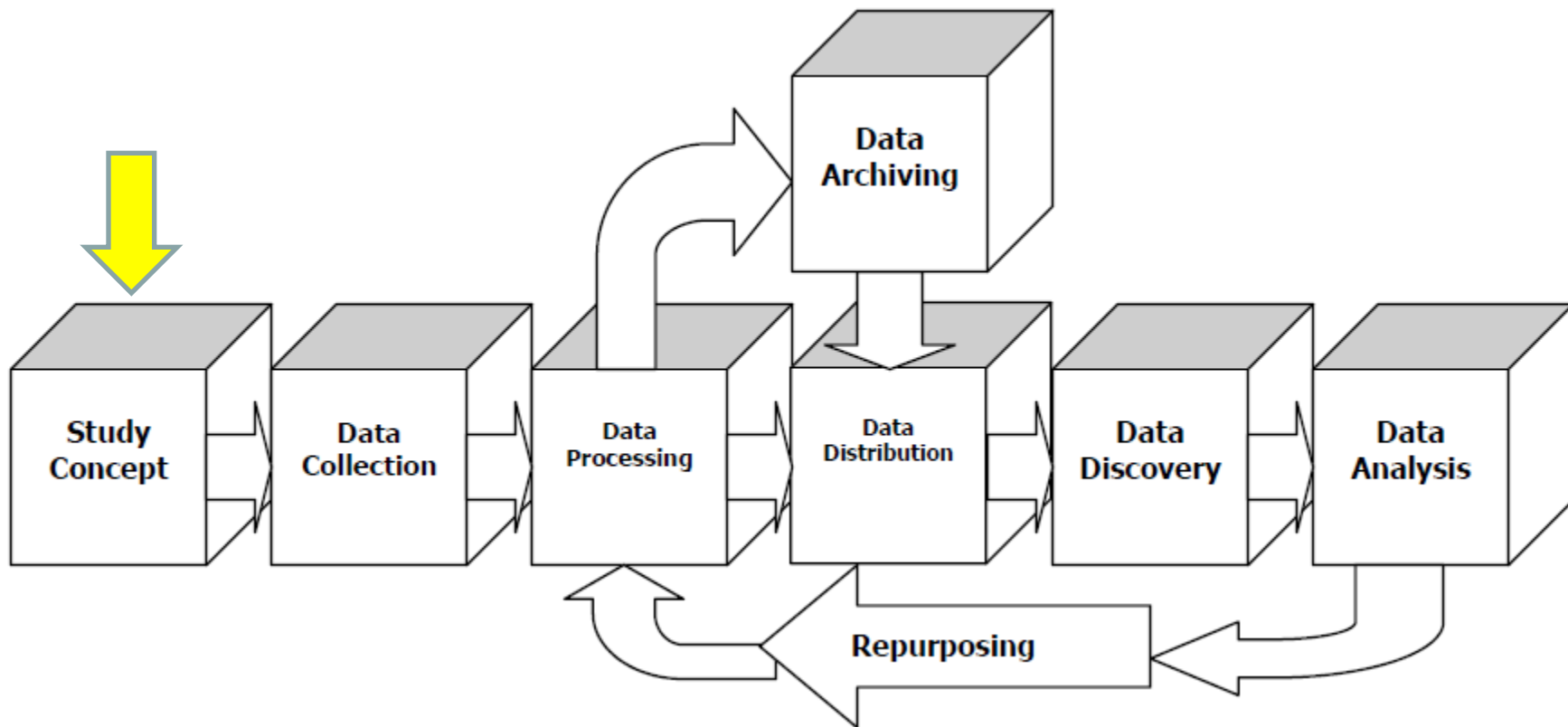
Data Documentation Initiative (DDI)

По версии 3.1

Сегодня покажем

- **способы минимизации смысловых ошибок на самой первой стадии цикла работ**
- **роль «экспертов предметных областей» и рекомендации по организации их работы**
- **способ стыковки их работы с работой инженеров по знаниям, системных аналитиков и архитекторов, проектировщиков данных**

В других моделях цикла информационных работ стадия работы с понятиями тоже первая. Но детализируется недостаточно.



Data Documentation Initiative (DDI)

По версии 3.1

Цикл информационных работ в «Инф. Пространстве»: Как определяется ядро метаинформации? (1, 20-10-2014)



«Вертикальные» и «горизонтальный» сегменты Инф. Пространства эл. региона (2004 – 2007 г.г.)

(2004
–
2007
г.г.)

Целевые пользователи ЭР: жители, представители предприятий, туристы и др.

Базовая ИКТ-инфраструктура

Социальные сети и сообщества жителей

Электронное правительство региона

Электронная культура

Электронные СМИ

Электронное образование

Электронная медицина

Электронная демократия

Электронный бизнес

Межотраслевые системы (включая общее управление процессами, информационные ресурсы)

Средства интеграции



(2006 – 2008 г.г.)

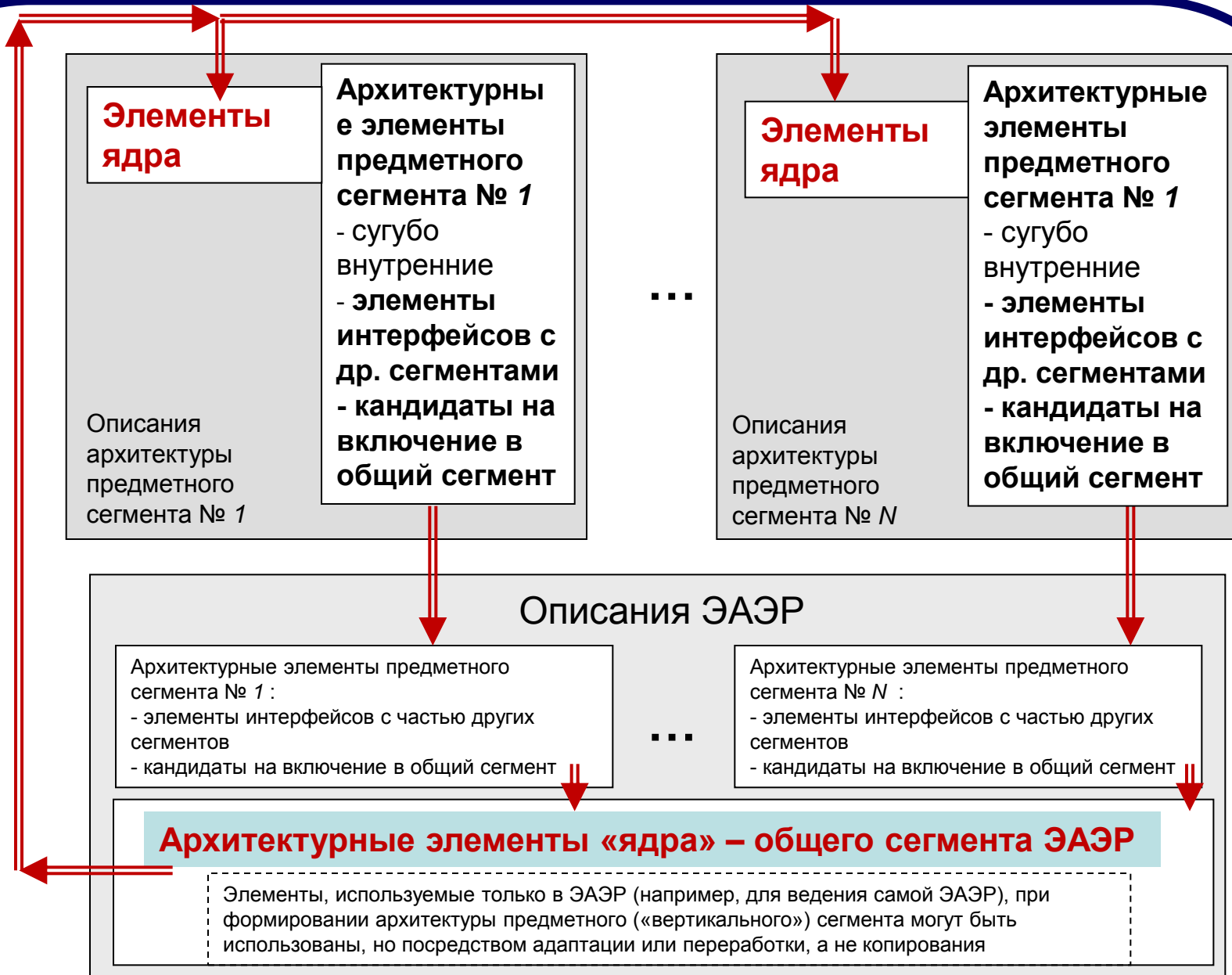
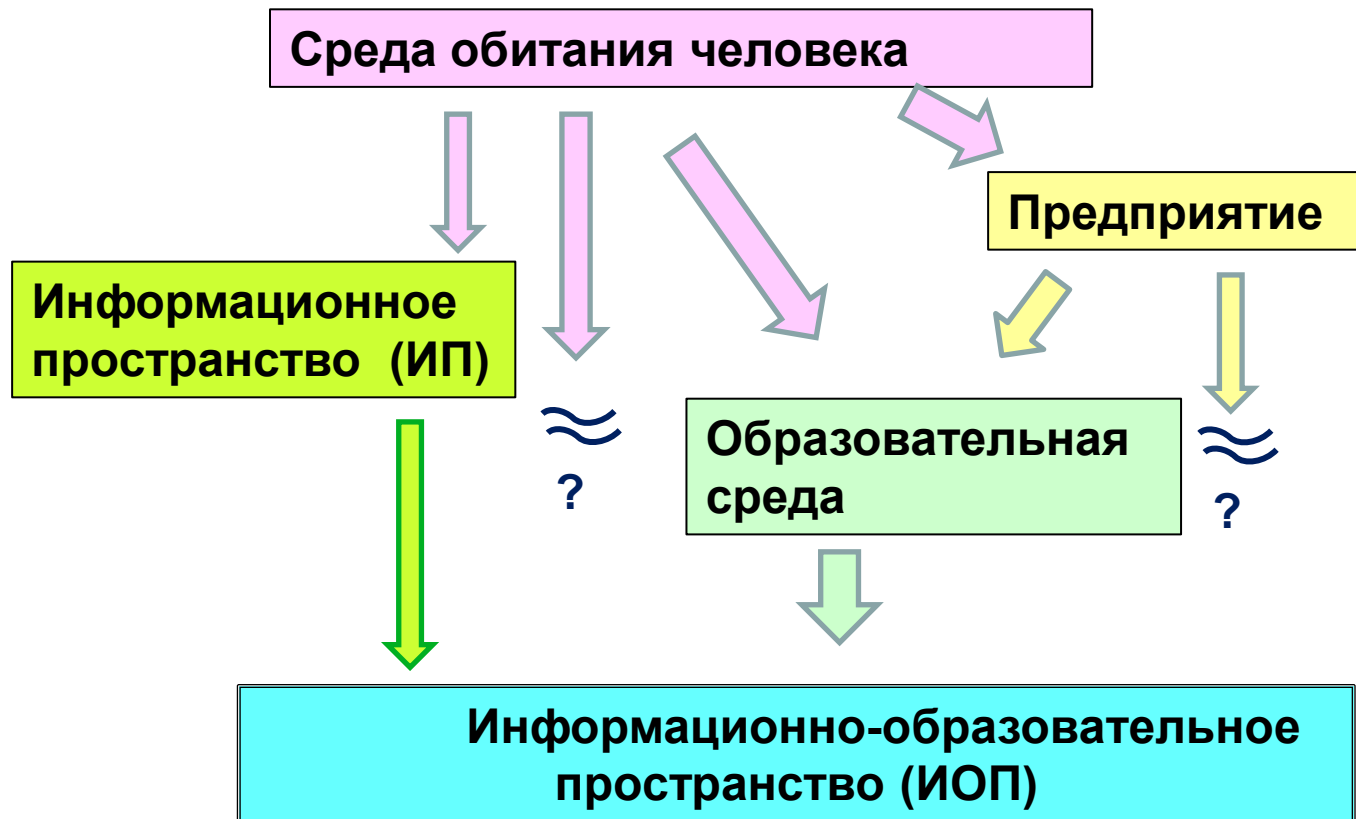


Рис. XX.XX. Взаимосвязи описаний Эталонной Архитектуры Эл. Региона и предметных сегментов электронного региона

Для ИОП -- Из проекта «Информационные пространства: требования к фундаментальным свойствам»



→ Наследование фундаментальных свойств высокоуровневых категорий и понятий - фундамент совместимости и надежности

Примеры смысловых несоответствий:

1) Что должно быть получено на запрос предоставить «сетевое обучение»:

- по закону «Об образовании в РФ» (N 273-ФЗ)
- по стандартам на ИТ в образовании (ГОСТ Р 53620, ГОСТ Р 52653)

2) По какой причине гражданин США не сможет заполнить форму – заявку на визу в РФ, эквивалентную форме DS-160 в вопросах о

- родителях
- местонахождении родственника

ПРОБЛЕМЫ и РЕШЕНИЯ связаны со КОЛЛИЗИЯМИ в СЕМАНТИКЕ

Проблемы не только для «Эл. региона», но и на уровне одного ведомства и его связей с другими ведомствами

(1994

–
1995
г.г.)

Пример: программа проектов ГУВД Москвы 1994-1995 г.г.: «**Системы ИО оперативно-розыскной деятельности, Единая модель данных**»

Корневые понятия и категории информации ЕМД:

- *Лицо*
- *Место*
- *Предмет*
- *Дата и время*
- *Событие*
- *Роль*
- *Связь*

Требования информации других ведомств (стрелки – примеры пересечения с другими ведомствами)

(1994
–
1995
г.г.)

Лицо

- Физическое → ЗАГС, Минтруд, ...
- Юридическое → Минюст, ФНС, ...

• Место

- Территория → Росреестр ...
- Адрес → ...
- Координаты → ...

• Предмет

- Оружие →
- Документ →
- Вещь ← Антиквариат ...
НО! Это и оружие
и документ
- Сооружение →

• Дата и время

- Момент
- Интервал

• Событие

- ...

• Деятельность

- ...

• Роль

- ...

• Связь

- ...

Проблемы с классификаторами, несовместимостью справочников, и т.д.

- **КАК ИЗБЕЖАТЬ ТИПОВЫХ ОШИБОК ПРИ ПОСТРОЕНИИ КОРПОРАТИВНЫХ И ОТРАСЛЕВЫХ СИСТЕМ НОРМАТИВНО-СПРАВОЧНОЙ ИНФОРМАЦИИ**

[Дмитрий Гулько](#) *PC Week/RE от 25.5.2004*

- «...Формально классификатор РФФИ представлен как чисто иерархический, хотя в нем **невооруженным глазом просматривается неоднозначность** (...разделы 01-326 *Динамика атмосферы и океана* перекликаются с разделами 05-513 и 05-514 в *океанологии* и с разделами 05-610 и 05-612 в *физике атмосферы*), ... эта неоднозначность напрямую связана с насильственной попыткой фасетную классификацию втиснуть в прокрустово ложе иерархической. »
(<http://serendip.narod.ru/order/classif/clas5.html>)

Что делается «мозгом вручную»?

Сбор и анализ информации об информации

Анкеты, Словари синонимов, омонимов...
Классификаторы, кодификаторы

НЕОДНОЗНАЧНЫЕ ПРЕОБРАЗОВАНИЯ

Фиксация понимания
– семантика

Глоссарии, Тезаурусы
Классификаторы («коды»)

НЕОДНОЗНАЧНЫЕ ПРЕОБРАЗОВАНИЯ

– формальное
представление
интегрированной, «единой»
семантики

Плоскости абстракции и агрегации
Разновидности семантических связей и их
модальности
Темпоральные связи
Семантическая близость и условные связи
Семантические сети, онтологии
RDF-схемы ?? Онтологии Gellish ??

НЕОДНОЗНАЧНЫЕ ПРЕОБРАЗОВАНИЯ

– логические структуры

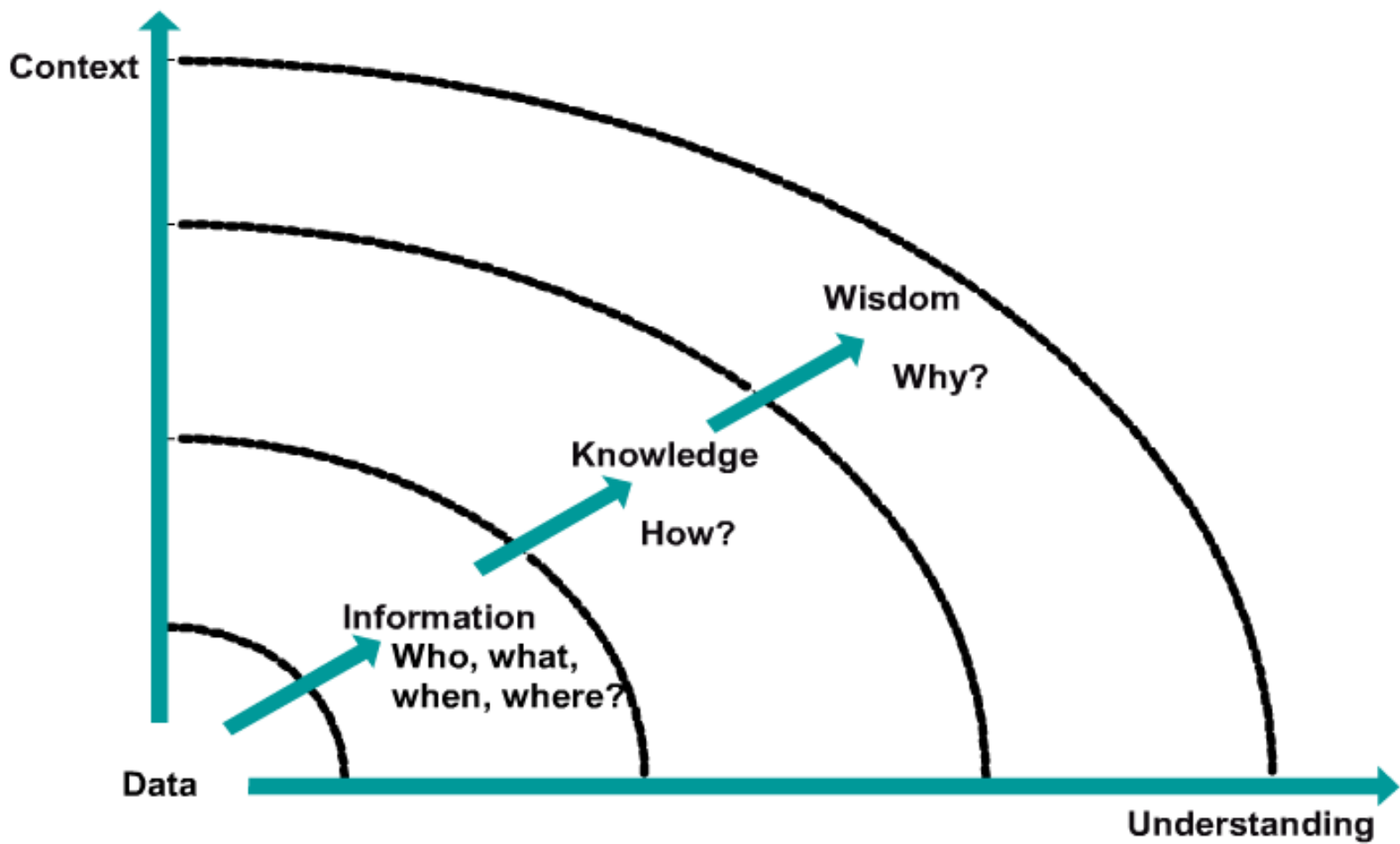
SQL-схемы, XML-схемы

NIEM NDR 9.2 Usage of English

The name of any XML Schema component defined by the schema SHALL be composed of words from the English language, using the prevalent U.S. spelling, as provided by **[OED]**.

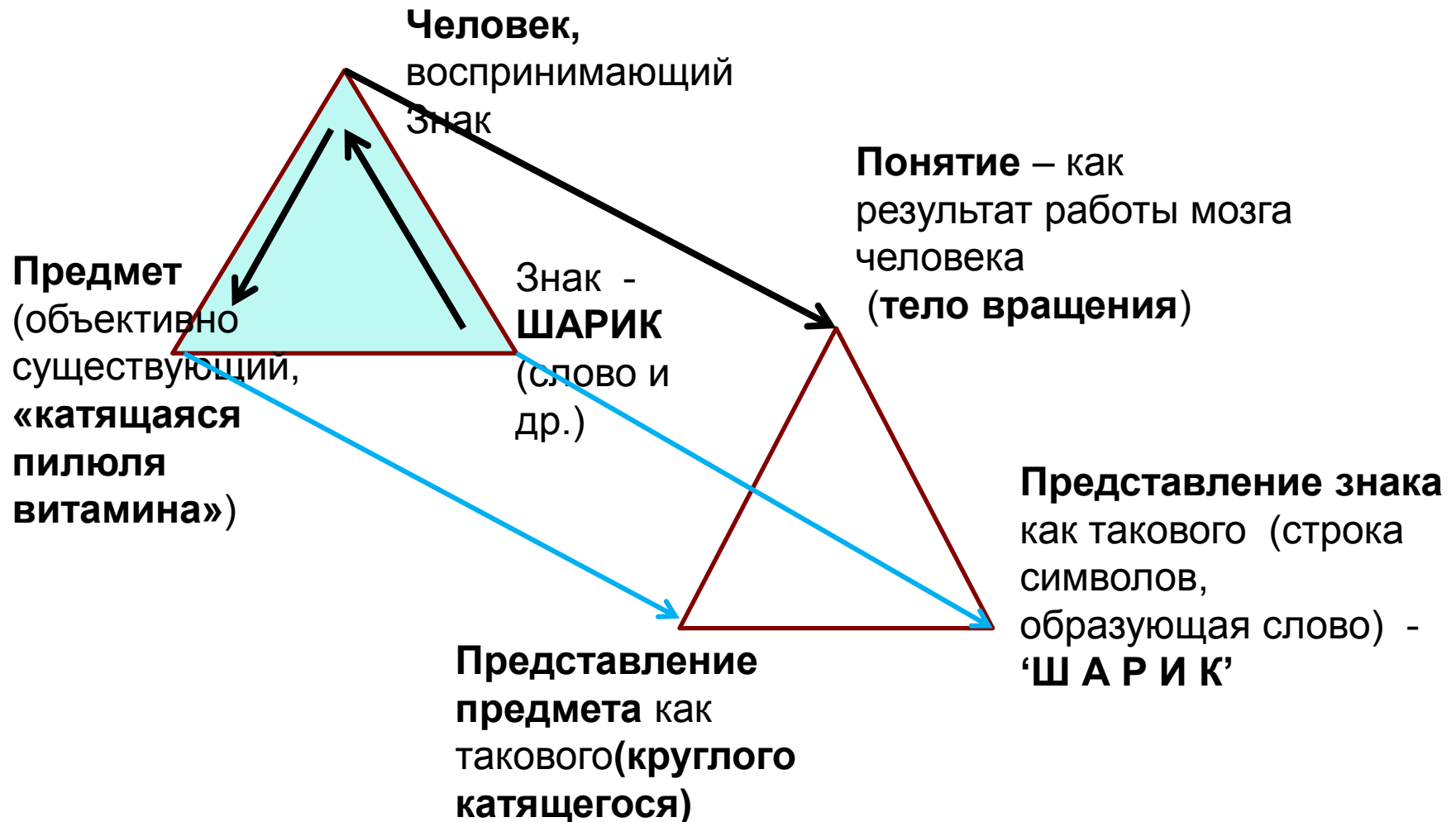
Rationale The English language has many spelling variations for the same word. For example, American English “program” has a corresponding British spelling “programme.” **This variation has the potential to cause interoperability problems** when XML components are exchanged because of the different names used by the same elements. Providing users with a dictionary standard for spelling will mitigate this potential interoperability issue.

«Путь к знаниям»



Проблемы

- отсутствия «единой информации»,
- размывания и искажения её со временем



Корни проблемы отсутствия «единой информации»

Человек №1 → понятие №1

(«образование, предоставляемое несколькими членами сети ОУ»)

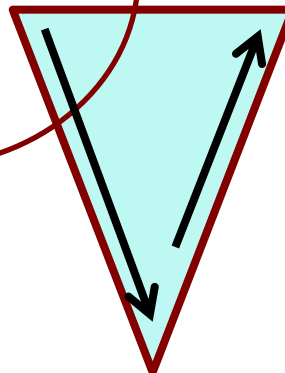
???

Предмет №1
(образование в МГУ и Сорбонне с двумя дипломами)



Знак -
«Сетевое обучение»

Предмет №2
(«образование, получаемое в МГУ, но дистанционно, с помощью ИКТ»)

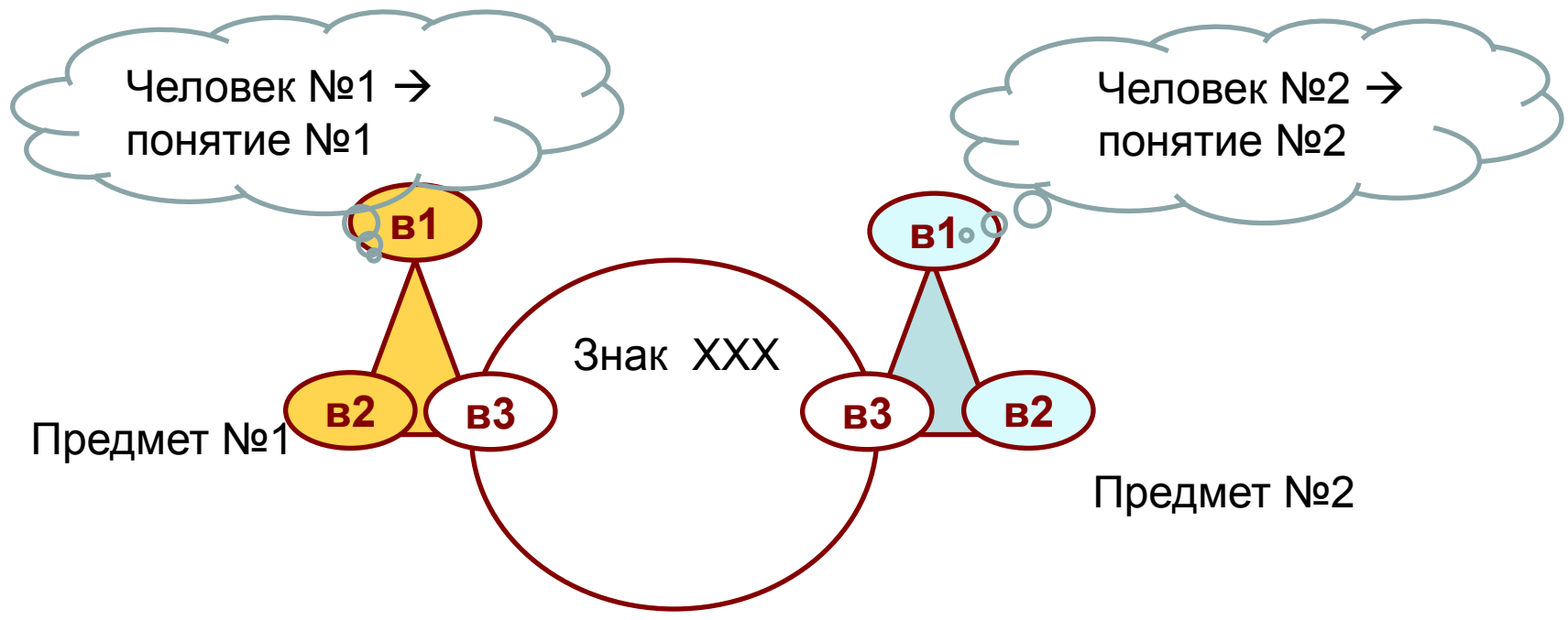


Человек №2 → понятие №2

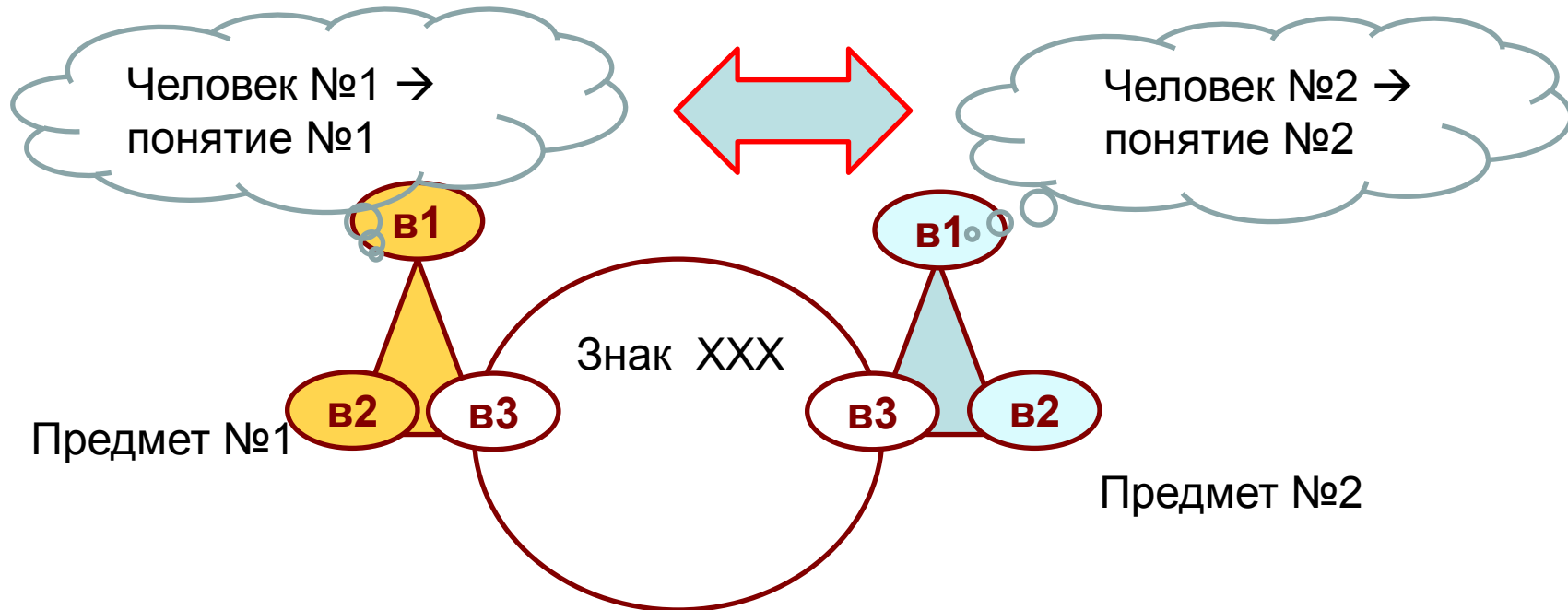
(«образование, получаемое в сети распределенной ИКТ-системы»)

???

Стартовые проблемы семантики –

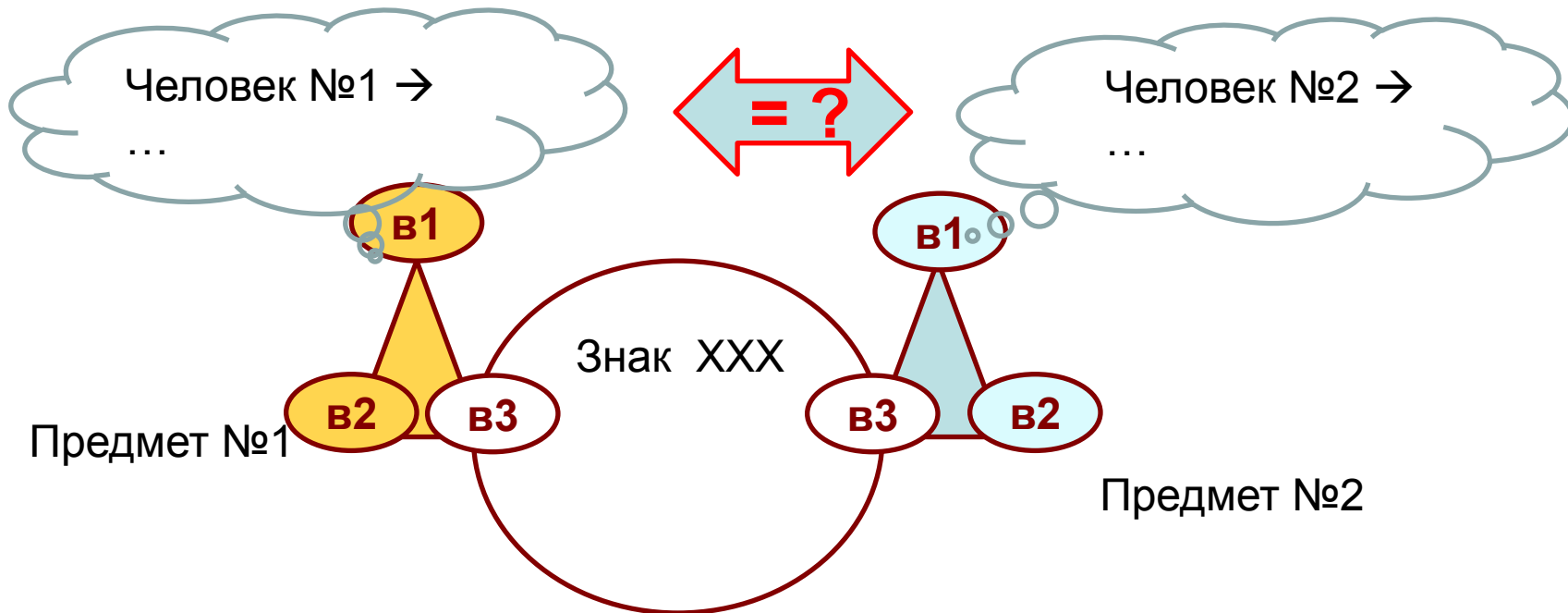


Стартовые проблемы семантики – - найти общее понимание Экспертов Предметных областей



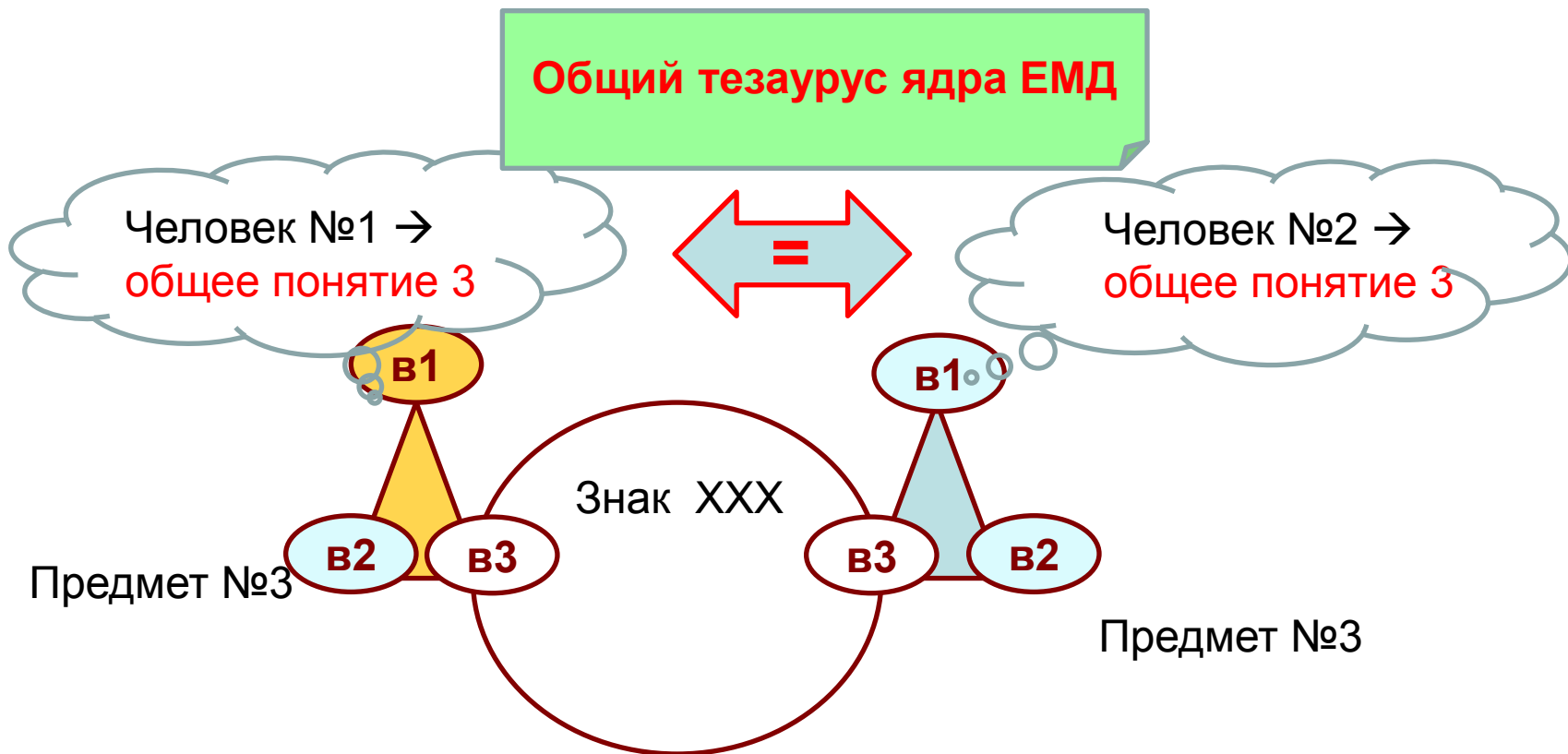
- 1) Переговоры о задачах и понятиях
- 2) Рабочие описания содержания и выделения характеристик понятий

Стартовые проблемы семантики – - найти общее понимание Экспертов Предметных областей



- 3) Нормализация категорий в плоскостях обобщения и агрегации
- 4) Корректные классификации

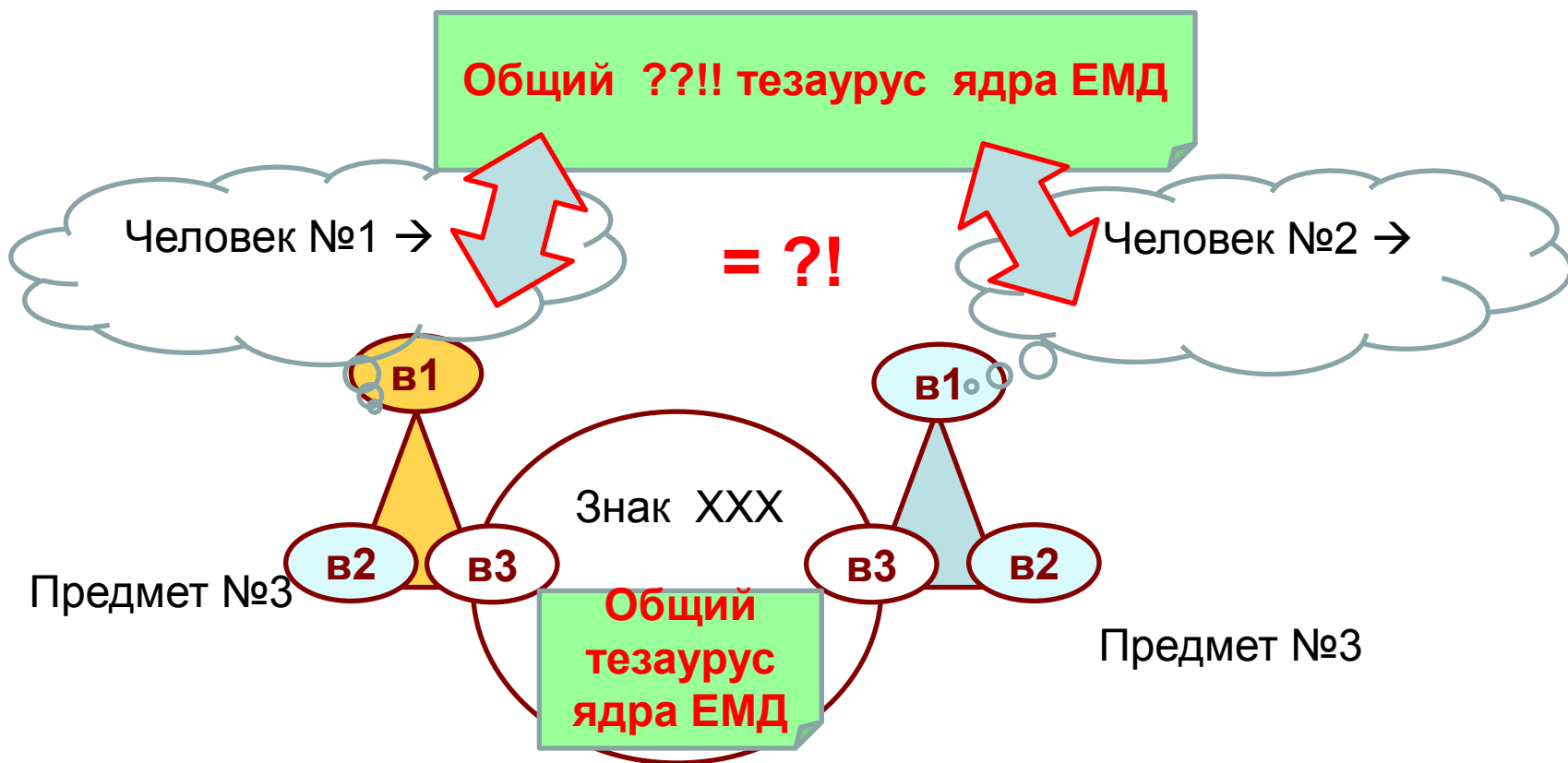
Стартовые проблемы семантики – - описать общее понимание Экспертов Предметных областей



5) Построение глоссариев и тезаурусов

6) Передача глоссариев и тезаурусов «гл. инженеру по знаниям» и достижение общего понимания с ним

Отложенные проблемы семантики – - потери / искажения общего понимания с течением времени и новыми Экспертами ПРО



7) Постоянный мониторинг сохранения семантики – изменение нормативно-правовой информации, технологий, практики, новые поколения специалистов и пользователей, и др.

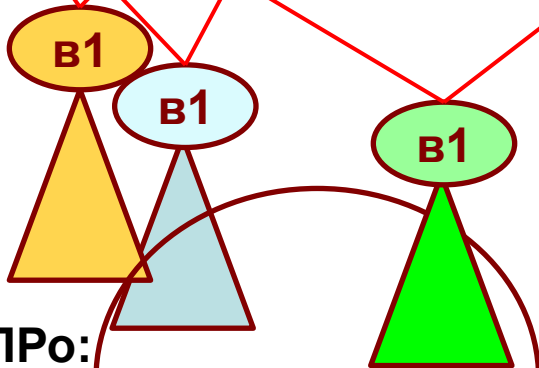
Коммуникация и общее понимание – формализованная, но общедоступная запись общего понимания

Общий тезаурус ядра ЕМД

«Полуформальное описание» --
Формализованная запись описания
понятия и правил его интерпретации

На естественном языке –
Запись описания понятий,
смысловых связей, правил интерпретации

Эксперты ПРО:
Человек №1
Человек №2

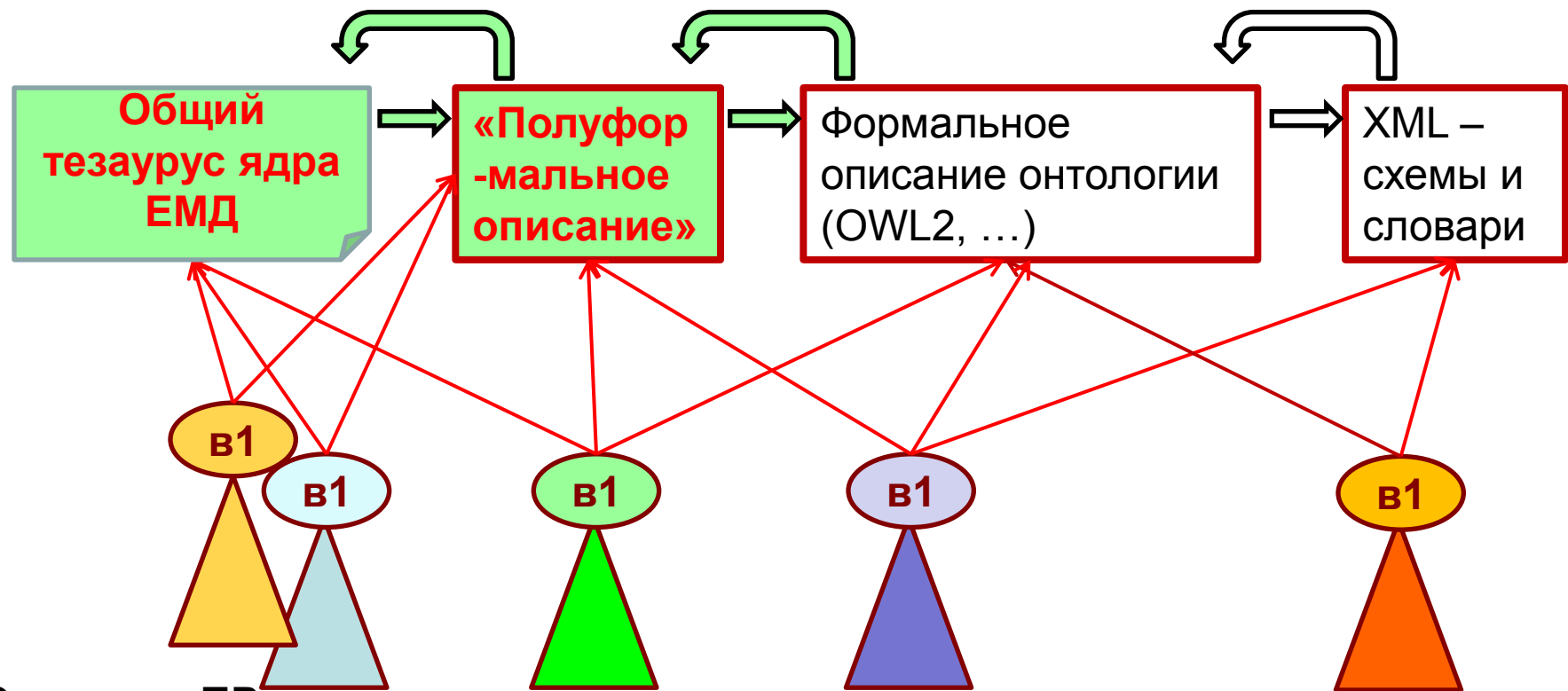


Знак XXX

Исходные сведения для передачи системным аналитикам и архитекторам для проверки и преобразований в форму, пригодную для интерпретации «ИТ-людьми»:
Простые (частичные) блок-схемы, диаграммы, решающие таблицы и др.

Человек №3 – «Ведущий координатор семантической интеграции ЕМД», «инженер по знаниям»

Коммуникация и общее понимание: формальная запись общего понимания и необходимость трассировки



Эксперты ПРО:
Человек №1
Человек №2

Человек №3 –
«Ведущий координатор семантической интеграции ЕМД», «инженер по знаниям»

Человек №4 –
«Системный аналитик / архитектор» (по инф. архитектуре)»

Человек №5 –
«Проектировщик схем данных ЕМД и интеграции данных сегментов

Отсюда следует

- 1) Необходимость определять Центральную группу экспертов ПрО (включая общих юристов) и ее роль
- 2) Необходимость давать им методику, процедуру и простой инструмент для самостоятельного (!) определения, фиксации и контроля семантики – смысла общей информации в ПрО
- 3) Поэтапная Формализация, но в форме, понятной им, юристам, инженерам по знаниям и ИТ
- 4) Итерации и еще раз итерации

Спасибо за внимание!

Ваши вопросы?

Зиндер Е. З., НО Фонд ФОСТАС

E-mail: ezinder@fostas.org

ДОПОЛНИТЕЛЬНЫЕ СЛАЙДЫ

Какого типа преобразования в ИП нужны?

Для каждого разумен и нужен свой набор средств

1) Распознавать «заголовки» и адресовать сообщения?

→ распознавать минимум элементов и их кодов, обеспечить общий набор отображаемых символов, оставить распознавание реципиенту-человеку

2) Распознавать «заголовки», адресовать сообщения, контролировать выполнение заявок по простым критериям?

→ м.б. распознавать большее число типов элементов (атомов) и их кодов, но все еще в пределах специфических расширений дублинского ядра

3) Распознавать все сообщения полностью и включать их элементы (атомы) во внутреннее ИП получающей организации?

- производить отображение не только «имён», но имен со всеми релевантными семантическими связями в схему получателя,
- проверять эквивалентность концептуальных схем и границы их смысловой совместимости,
- корректно преобразовывать структуру данных для помещения в «свою» схему без потери информации

Если для ИШ ЭП принять достаточность:

2) Распознавать «заголовки», адресовать сообщения, контролировать выполнение заявок по простым критериям?

→ м.б. распознавать большее число типов элементов (атомов) и их кодов, но все еще в пределах специфических расширений дублинского ядра

то и в этом случае

- «ручная работа» все еще требуется на первых шагах этапа анализа и формирования семантики: анализ, фиксация понимания, формальное представление
- и для следующих шагов рекомендуется ограничивать схемы онтологий упрощенными графами (например, без транзитивных связей), допускающими простые проверки соответствий и эквивалентности.

Системы с неоднородными компонентами

НАДО ОПРЕДЕЛЯТЬСЯ С ТЕМ, КАКИЕ СИСТЕМЫ РАССМАТРИВАЮТСЯ:

- Типа корпоративных ИС или
- т.н. Информационные пространства (ИП)

Качественный переход с уровня корпоративных систем (шире объединения, отрасли и т.д.):

- Наличие определенного центрального субъекта (как сообщества) или **отсутствие**
- С поддержкой «Среды обитания» -- одной или нескольких и каких

и другие качественные отличия.

«Корпоративный» подход (до 2007 года)

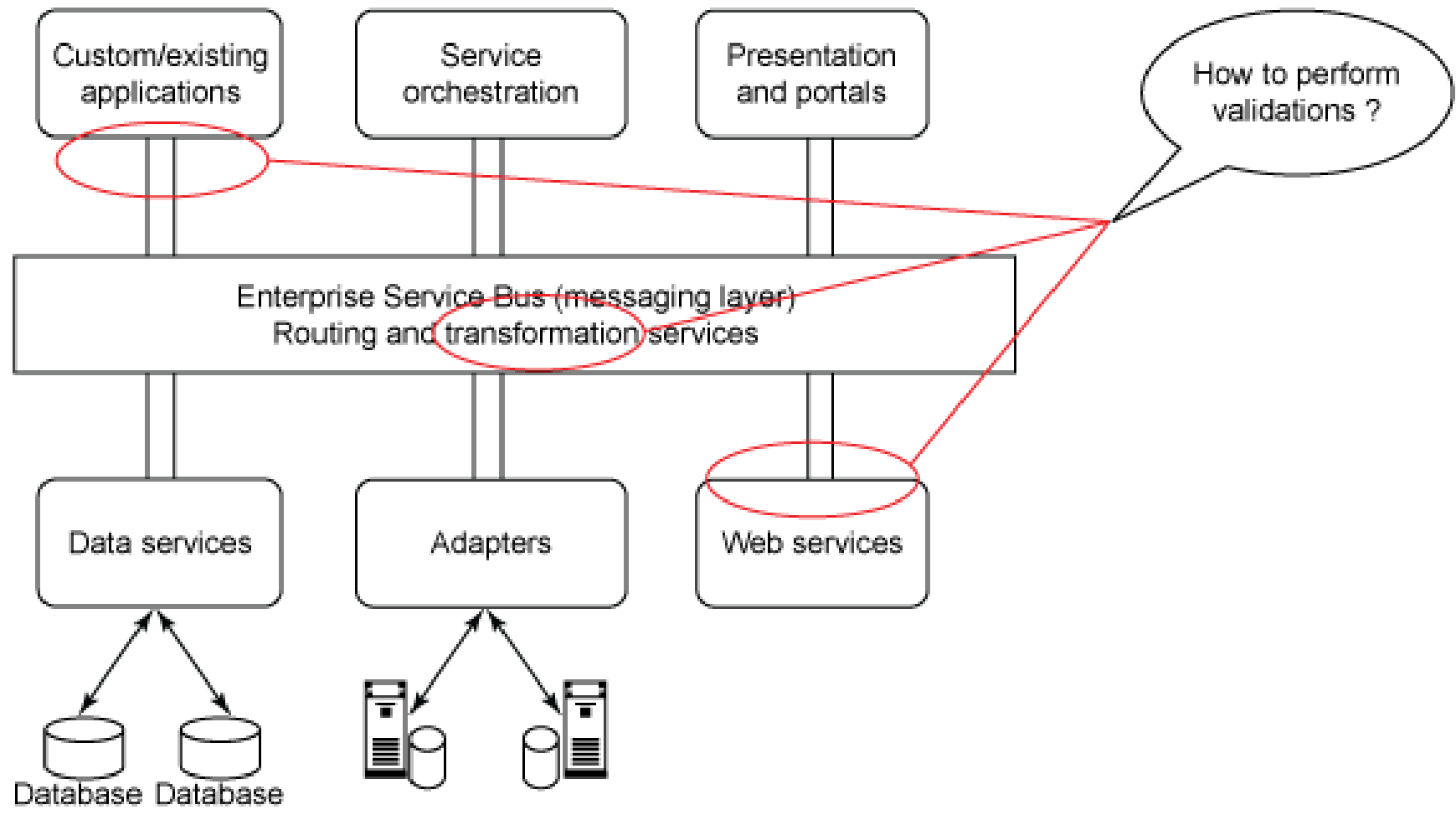
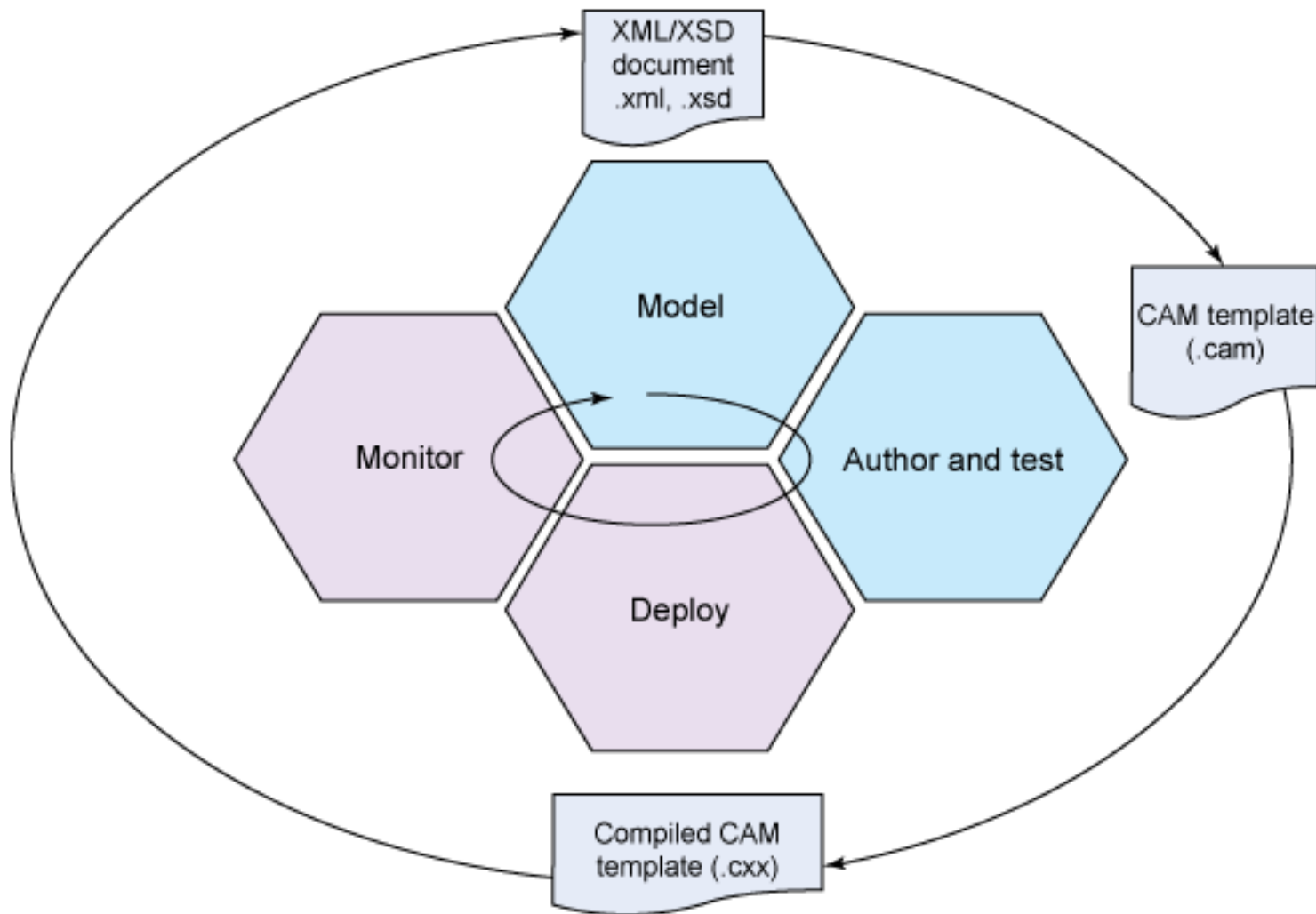


Схема проверок (и возможных преобразований) в САМ от IBM: постоянство и тяга к централизации



Модели, инструменты и организация

Нужное в проектировании ИБД и ИП известно после 1975:

- **типы моделей** (т.к. компьютерные семантические сети разработаны ранее)
- **как и типы переходов между моделями.**

К сожалению:

- виды моделей и переходов И ДАЖЕ «что такое семантика» часто определяют не сутью, а инструментами (((
- инструменты (языки в т.ч.) и процедуры имеют ограничения, часто непреодолимые (NP-полнота, ограничения на виды семантических связей, ...)

Инструменты (включая эффективные преобразователи, “Data Handlers”) **есть не для всех моделей и переходов**, что-то пишется at hoc

Без организационных структур и усилий

- не будут построены модели (и в Wikipedia есть организация!)
- не будут заполнены позиции в процессе взаимодействий и интеграции, где нет подходящего инструмента