

Организация проектно-аналитической работы при создании информационных систем

Лекция 2

Часовских А.А.

Говорко А.А.

МОДЕЛИ и НОТАЦИИ

Модели и нотации

Определение 1.

Под *моделью* понимается полное, точное и адекватное формальное описание предметной области (системы), имеющее конкретное назначение.

Описание является моделью системы, если оно может быть использовано для получения ответов на вопросы относительно системы с некоторой наперед заданной точностью.

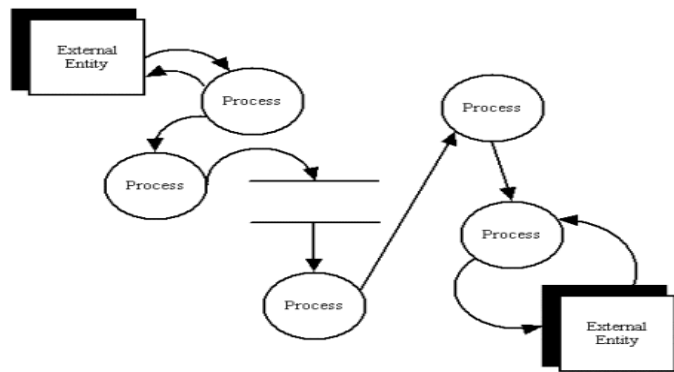
Ключевые параметры модели:

- Назначение модели (purpose)
- Совокупность вопросов, на которые должна отвечать модель (фактически, детализирующая назначение модели)
- Границы предметной области (scope)
- Точка зрения (viewpoint)
- Точность (как правило, неявно подразумевается)

Модели и нотации

Определение 2.

Нотация - система условных обозначений, состоящая из множества символов, используемых для представления понятий и их взаимоотношений (алфавит нотации), а также правил их применения.

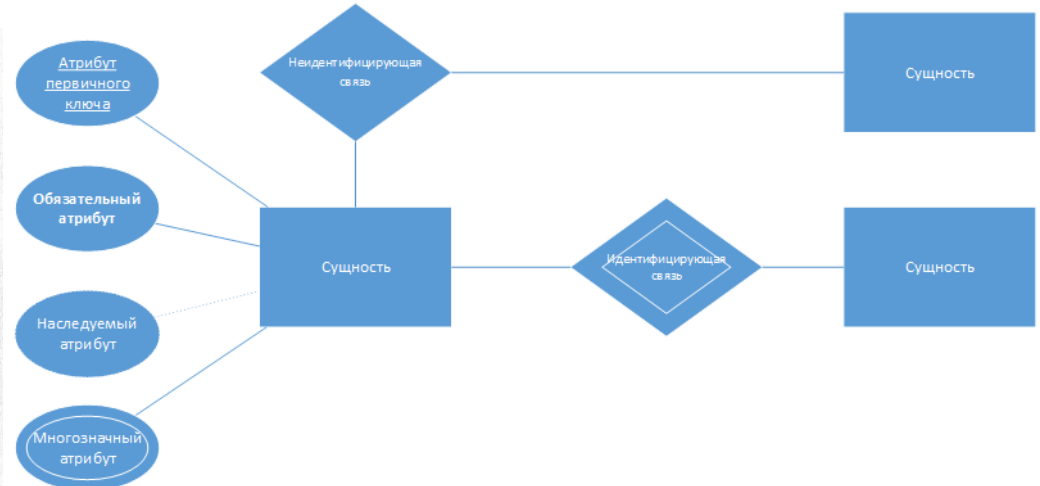
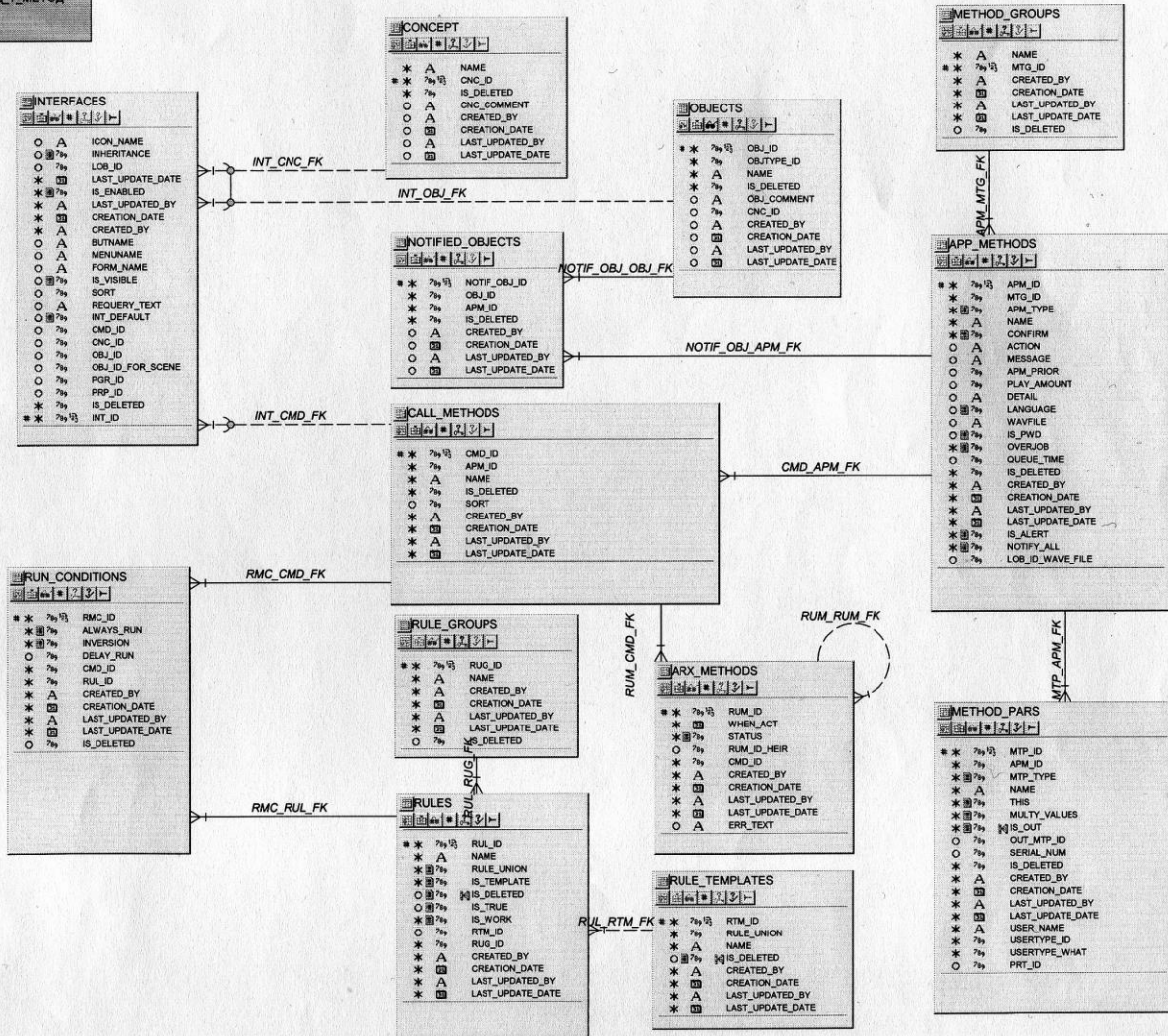


Пример
«диаграммы потоков данных» (DFD)

	Нотации Йордана (Yourdon), Де Марко (De Marco)	Нотация Гейна-Сарсона (Gane & Sarson)	Нотация SSADM	«Unified»-нотации (UML и др.)	BPMN
Интерфейс к внешней сущности					
Процесс				<code><<process>></code>	
Хранилище данных (или объект)				<code><<entity>></code>	
Поток (данных)					

Примеры диаграмм «сущность-связь»

Diagram: СДТ_МЕТОД
Title: Метод



Виды моделей

Основные виды моделей (в структурном проектировании):

- модель процессов (или функциональная модель)
- модель данных (или информационная модель)

Некоторые методологии могут вводить другие модели, например,

- диаграммы использования (Use Case)
- диаграммы состояний

Сбор данных (извлечение знаний)

Источники:

- Документы
- Наблюдение
- Анкетирование
- Собственные знания
- Интервью
- Придуманные описания

Точка зрения:
предметная область

Эксперт

Язык:
естественный



Точка зрения:
предметная область

Аналитик

Язык:
*естественный /
формальный*

Типы опросов (зачем?):

- Факты и проблемы
- Принятие решений
- Элемент цикла автор/читатель

Процесс опроса

1. Подготовка

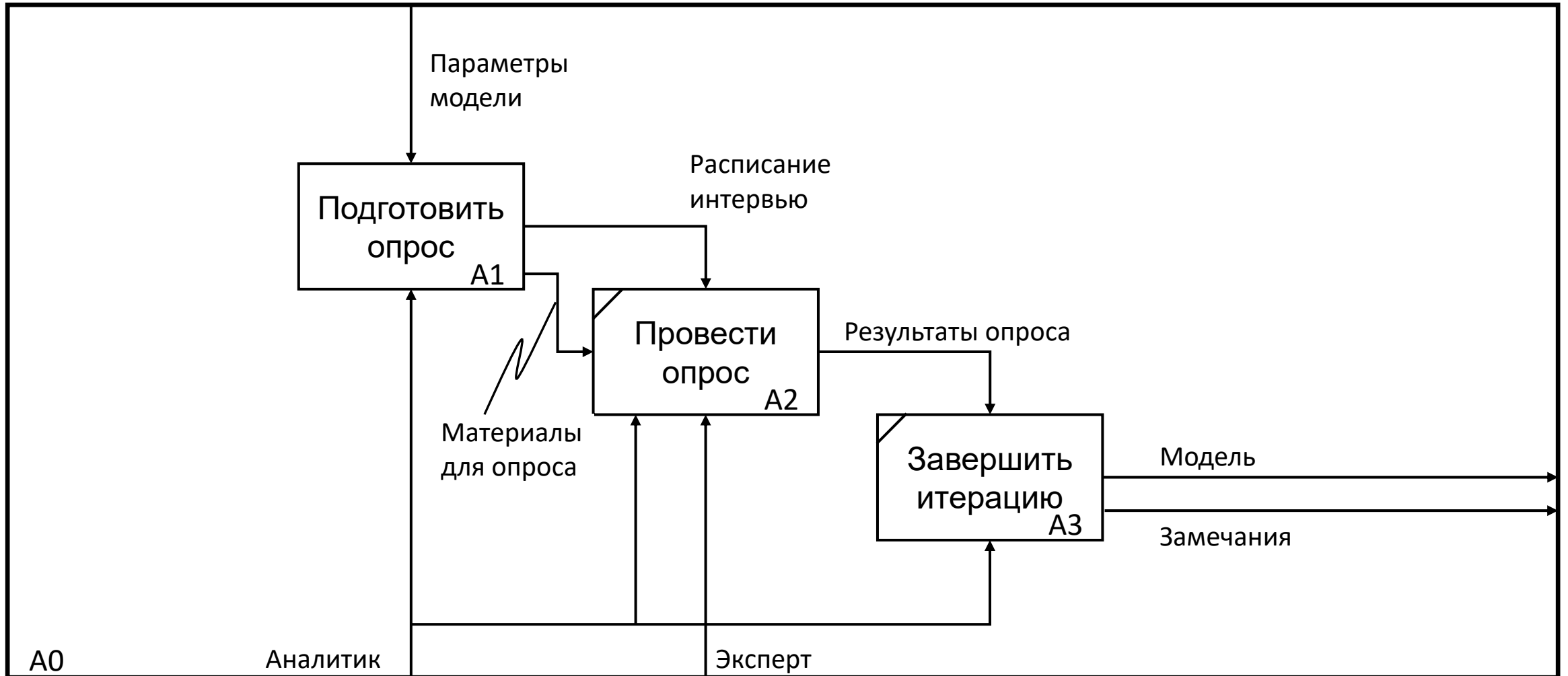
- выбор эксперта (экспертов)
- подготовка аналитика
 - темы / вопросы для обсуждения
 - сбор предварительной информации
- организационные вопросы
 - согласование в проектной группе
 - назначение встречи
- подготовка эксперта

2. Проведение

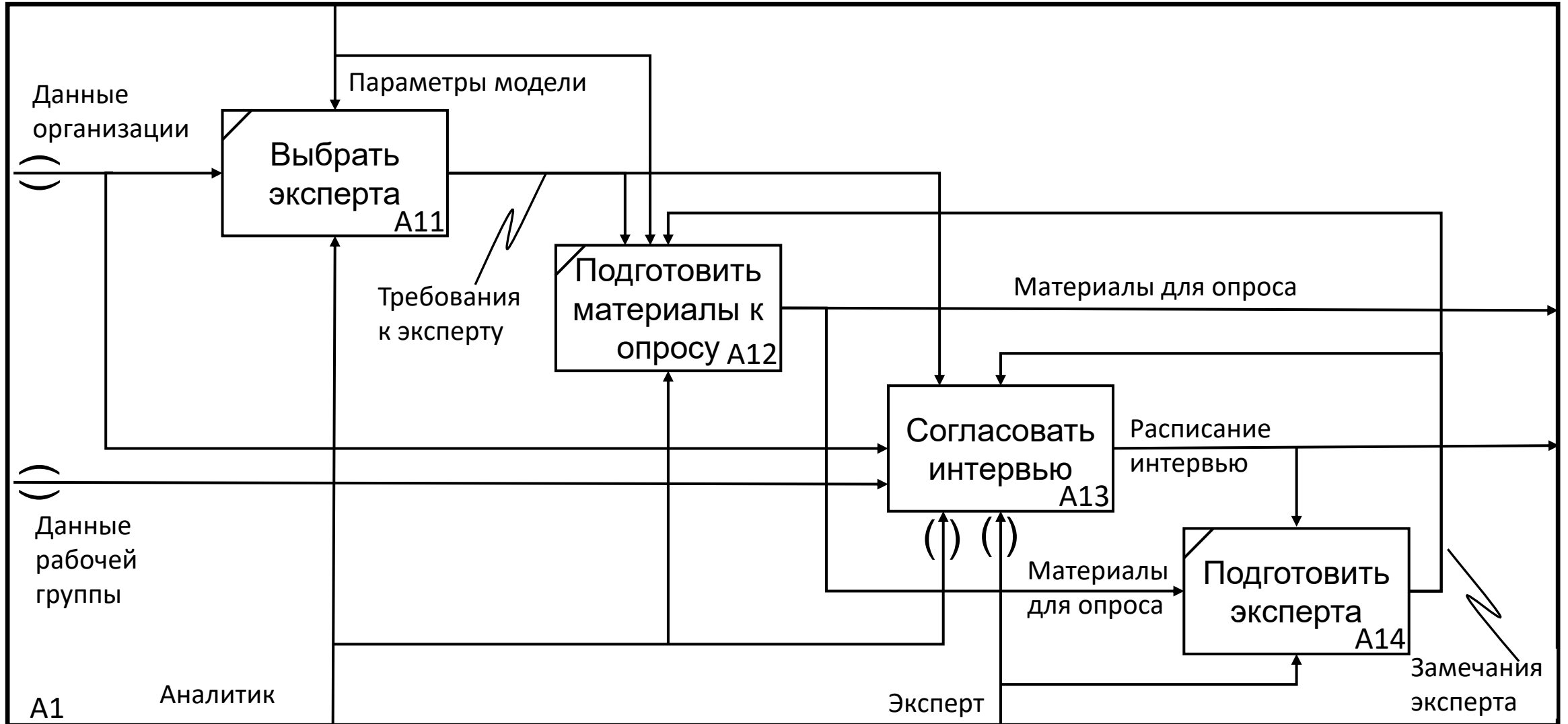
3. Завершение

IDEF-0

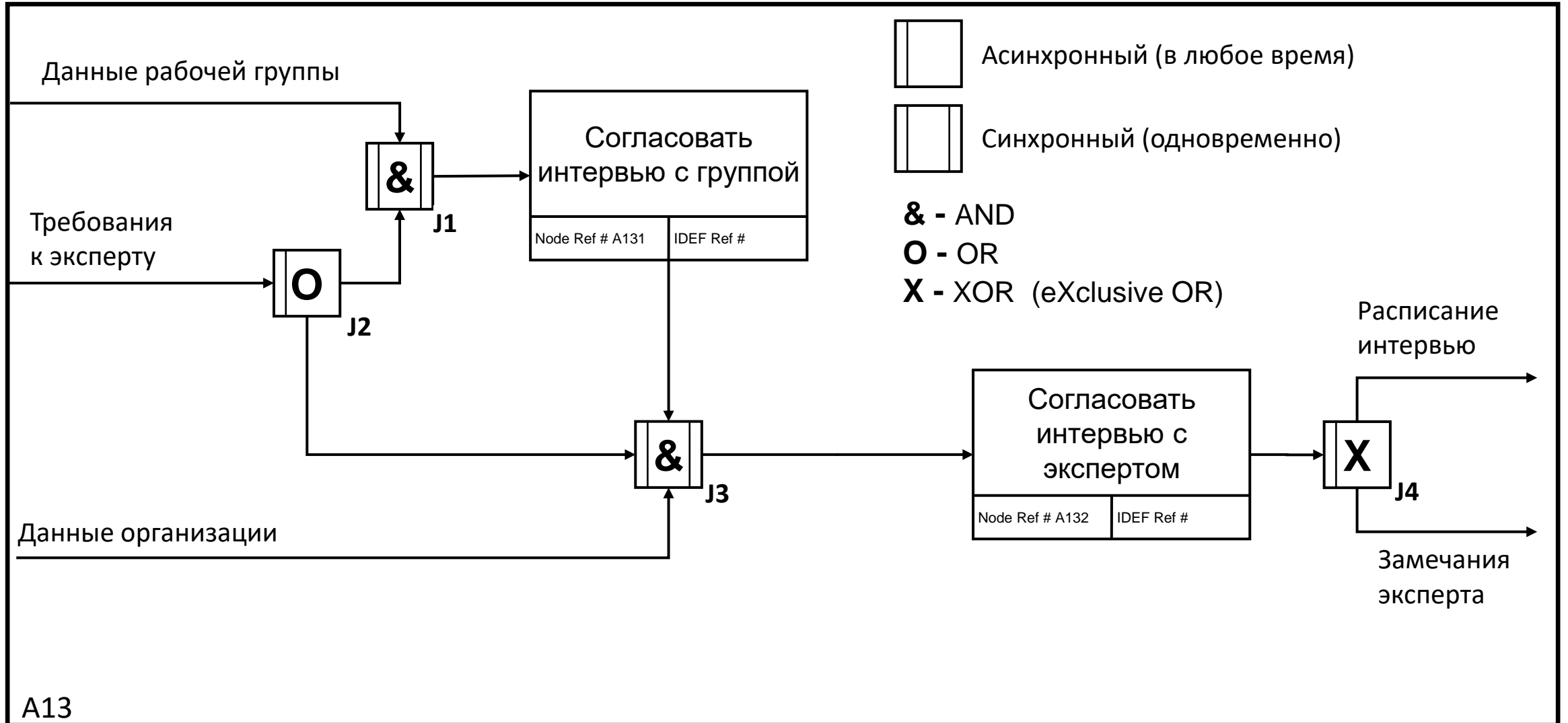
Модель процессов (1/3)



Модель процессов (2/3)



Модель процессов (3/3)



МОДЕЛИ ДАННЫХ

Отношения (1/2)

Определение 0. Домен D - абстрактный тип данных, множество «однородных» допустимых значений (эмпирическое) (и возможные операции).

Определение 1. Множество доменов $\{D_1, D_2, \dots, D_m\}$ ($m \geq 1$). Отношение $E \subseteq (D_1 \times D_2 \times \dots \times D_n)$ ($n \geq 1$).

Определение 1'. Отношение $E = \langle E^H ; E^B \rangle$. Заголовок $E^H = \{\langle A_i, D_i \rangle, i=1, \dots, n\}$, A_i - имя атрибута, D_i - домен атрибута, $A_i \neq A_q$ для любых $i, q=1, \dots, n$. Тело отношения $E^B = \{\langle A_1 : x_{1j}, \dots, A_n : x_{nj} \rangle\}$, $j=1, \dots, k$, $\forall i, j x_{ij} \in D_i$

A_1	A_2	...	A_n
x_{11}	x_{21}	...	x_{n1}
x_{12}	x_{22}	...	x_{n2}
...
x_{1k}	x_{2k}	...	x_{nk}

n атрибутов (полей таблицы) – заголовок отношения

k уникальных n -местных кортежей (строк таблицы) – тело отношения

Определение 1''. Отношение E - предикат вида $E: D_1 \times D_2 \times \dots \times D_n \rightarrow \{0;1\}$.
Кортеж $X = \langle x_1, \dots, x_n \rangle$ принадлежит E тогда и только тогда, когда $E(X) = 1$.

Отношения (2/2)

Определение 2. Пусть $A^1 = \{a_i^1\}$ и $A^2 = \{a_i^2\}$ - два подмножества множества A атрибутов отношения E . A^2 функционально зависит от A^1 (записывается $A^1 \rightarrow A^2$), если для любых кортежей $X = \langle x_i \rangle$ и $Y = \langle y_i \rangle$ отношения E таких, что $\langle a_i^1, x_i \rangle = \langle a_i^1, y_i \rangle$ для каждого i , верно $\langle a_j^2, x_j \rangle = \langle a_j^2, y_j \rangle$ для любого j .

В этом случае A^1 иногда называют детерминантом функциональной зависимости.

Определение 2. Если $A^1 \rightarrow A$, и ни одно собственное подмножество A^1 таким свойством не обладает, то A^1 - потенциальный ключ.

A_1	A_2	A_3	A_4	A_5	...	A_{n-1}	A_n
x_{11}	x_{21}	x_{31}	x_{41}	x_{51}	...	x_{n1}	x_{n1}
x_{12}	x_{22}	x_{32}	x_{42}	x_{52}	...	x_{n2}	x_{n2}
...
x_{1k}	x_{2k}	x_{3k}	x_{4k}	x_{5k}	...	x_{nk}	x_{nk}

$$A^1 = \{A_1; A_2\}, A^2 = \{A_3; A_4; A_5\}$$

$$(x_{11}; x_{21}) = (x_{12}; x_{22}) \Rightarrow (x_{31}; x_{41}; x_{51}) = (x_{32}; x_{42}; x_{52})$$

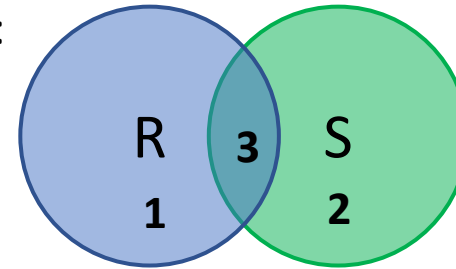
Операции над отношениями (1/3)

Операции реляционной алгебры – ок. 1970г. Эдгар Кодд (Edgar Frank "Ted" Codd, 19.08.1923-18.04.2003)

- Операции, аналогичные операциям над множествами:

- объединение,
- пересечение,
- вычитание (разность),
- декартово произведение;

$$R(r_1, \dots, r_n) \otimes S(s_1, \dots, s_m) = Z(r_1, \dots, r_n, s_1, \dots, s_m)$$



К. Дж. Дейт
«Введение в системы
баз данных»

- Специальные реляционные операции:

- выборка, $\sigma_{a\theta b}(R)$ или $\sigma_{a\theta c}(R)$, где $a, b \in \{r_1, \dots, r_n\}$, $c = const$
- проекция, $R[A]$, где $A \subseteq \{r_1, \dots, r_n\}$
- соединение, $R[X\Theta Y]S$, где $X \subseteq \{r_1, \dots, r_n\}$, $Y \subseteq \{s_1, \dots, s_m\}$, все заголовки и кортежи, для которых $X\Theta Y$. $R \bowtie S$ – заголовок $R \cup S$.

- деление. Отношения $R(r_1, \dots, r_n, s_1, \dots, s_m)$ и $S(s_1, \dots, s_m)$. Заголовок $R \div S = (r_1, \dots, r_n)$. Тело $R \div S = \{ \langle r_1, \dots, r_n \rangle \mid \forall \langle s_1, \dots, s_m \rangle \in S \exists \langle r_1, \dots, r_n, s_1, \dots, s_m \rangle \in R \}$.

$$(R \otimes S) \div S = R, \text{ однако } (R \div S) \otimes S \subseteq R.$$

Операции над отношениями (2/3)

Операторы соединения, пересечения и деления можно выразить через другие реляционные операторы:

- Оператор **соединения** определяется через операторы декартового произведения и выборки. Для оператора естественного соединения добавляется оператор проекции.
- Оператор **пересечения** выражается через вычитание следующим образом: $A - (A - B)$.
- Оператор **деления** выражается через операторы вычитания, декартового произведения и проекции следующим образом: $A[X \div Y]B = A[X] - ((A[X] \otimes B) - A)[X]$. Или $A[X] - ((A[X] \otimes B[Y]) - C[X, Y])[X]$.

Пусть d_a и d_b - степени отношений A и B .

Результат деления A на B есть множество кортежей p степени $(d_a - d_b)$ таких, что для всех кортежей $u \in B$ степени d_b кортеж $pu \in A$.

Обозначим через P проекцию отношения A на первые $(d_a - d_b)$ атрибутов.

Тогда $(P \otimes B) - A$ соответствует кортежам $r \notin A$, появившимся в результате конкатенации P с каждой строкой из B . Пусть V обозначает множество кортежей $p \in P$ степени $(d_a - d_b)$. Тогда $A[X \div Y]B = P - V$.

Оставшиеся операторы (объединение, вычитание, декартово произведение, выборка, проекция) - примитивные:

- **Декартово произведение** увеличивает количество атрибутов. Это единственный такой оператор.
- **Проекция** уменьшает количество атрибутов. Это единственный такой оператор.

Операции над отношениями (2/3)

- **Объединение** - единственный, кроме произведения, оператор, который увеличивает количество кортежей, но произведение также увеличивает количество атрибутов.
Для объединения A и B они должны быть совместимы по типу и в их объединении должны быть те же атрибуты, что и в них самих.
Если применить к произведению $A \otimes B$ проекцию, чтобы сократить множество атрибутов до множества атрибутов A (или B), то получится исходное отношение A (или B). (Если B - не пустое). Т.е. произведение также не может быть использовано.
- **Вычитание** не может быть выражено через произведение или объединение, поскольку они увеличивают количество кортежей, или через проекцию, поскольку она уменьшает количество атрибутов. Выборка - одноместная операция.
- Оператор **выборки** - единственный оператор, позволяющий проводить сравнения по атрибутам отношения, поэтому его нельзя выразить через объединение, вычитание, декартово произведение, проекцию.

Преобразование отношений (1/3)

Правила вывода Армстронга

Пусть A, B, C - подмножества множества атрибутов отношения E . Далее AB будет обозначать $A \cup B$.

- правило рефлексивности: если $B \subseteq A$, то $A \rightarrow B$;
- правило дополнения: если $A \rightarrow B$, то $AC \rightarrow BC$;
- правило транзитивности: если $A \rightarrow B$ и $B \rightarrow C$, то $A \rightarrow C$.

Декомпозиция

Отношение E с множеством функциональных зависимостей S декомпозируется на отношения E_i с множествами функциональных зависимостей S_i ($i=1, \dots, n$).

Декомпозиция (без потерь): $E_1 \bowtie E_2 \bowtie \dots \bowtie E_n = E$.

Сохранение функциональных зависимостей - равенство замыканий $S^+ = S'^+$, где S' - объединение множеств зависимостей S_i (в общем случае $S \neq S'$).

Преобразование отношений (2/3)

Теорема Хита (Heath)

Пусть $E(A,B,C)$ - отношение и A, B, C - подмножества его атрибутов. Если E удовлетворяет функциональной зависимости $A \rightarrow B$, то E можно представить как соединение проекций $E_1(A,B)$ и $E_2(A,C)$.

Обозначим (a,b,c) кортеж вида $\langle A:a, B:b, C:c \rangle$.

- 1) Ни один кортеж не утрачивается (при разбиении на проекции и обратном их соединении):
пусть $(a, b, c) \in E$, тогда $(a, b) \in E_1$ и $(a, c) \in E_2$, следовательно $(a, b, c) \in E_1 \bowtie E_2$.
- 2) Каждый кортеж соединения - кортеж из E ("не появилось новых"): пусть $(a, b, c) \in E_1 \bowtie E_2$.
Тогда для генерации кортежа соединения должны выполняться условия $(a, b) \in E_1$ и $(a, c) \in E_2$.
Для генерации кортежа $(a, c) \in E_2$ для некоторого атрибута b^* должен существовать кортеж $(a, b^*, c) \in E$.
Тогда также должен существовать кортеж $(a, b^*) \in E_1$.
Учитывая, что $A \rightarrow B$, из $(a, b) \in E_1$ и $(a, b^*) \in E_1$ следует, что $b = b^*$, следовательно $(a, b, c) \in E$.

Преобразование отношений (3/3)

Определение. E - отношение и A, B, C - произвольные подмножества множества его атрибутов. Подмножество B многозначно зависит от подмножества A (обозначается $A \twoheadrightarrow B$) тогда и только тогда, когда множество значений B , соответствующее заданной паре $(A:a, C:c)$, зависит от a , но не зависит от c . Многозначная зависимость $A \twoheadrightarrow B$ называется тривиальной, если выполнено хотя бы одно из условий « B является подмножеством A » или «объединение A и B образует все множество атрибутов E ».

Лемма Фейгина (о связанных парах). Для отношения $E(A, B, C)$ многозначная зависимость $A \twoheadrightarrow B$ выполняется тогда и только тогда, когда выполняется $A \twoheadrightarrow C$. Символически связанные пары многозначных зависимостей обозначают $A \twoheadrightarrow B | C$.

Теорема Фейгина (Fagin). $E(A, B, C)$ - отношение и A, B, C - множества его атрибутов. E равно соединению проекций $E_1[A, B]$ и $E_2[A, C]$ тогда и только тогда, когда $A \twoheadrightarrow B | C$.

Нормальные формы (1/2)

- Определение 1.** Отношение находится в 1НФ, если значения всех атрибутов атомарны.
- Определение 2.** Отношение находится в 2НФ, если оно находится в 1НФ и каждый неключевой атрибут неприводимо зависит от первичного ключа.
- Определение 3.** Отношение находится в 3НФ, если оно находится во 2НФ и каждый неключевой атрибут нетранзитивно зависит от первичного ключа (за исключением зависимостей, определяемых альтернативными ключами).
- Определение 3'.** Отношение находится в НФБК (Бойса-Кодда, Boyce/Codd), тогда и только тогда, когда каждая его нетривиальная (тривиальная - слева супермножество правой части) и неприводимая слева функциональная зависимость имеет в качестве детерминанта потенциальный ключ (т.е. детерминанты всех функциональных зависимостей - потенциальные ключи).

Нормальные формы (2/2)

Определение 4. Для отношения, находящегося в 4НФ, если существуют подмножества атрибутов A и B , такие, что выполняется нетривиальная многозначная зависимость $A \twoheadrightarrow B$, то все атрибуты функционально зависят от A . Иначе говоря, отношение находится в 4НФ, если оно находится в НФБК и все [многозначные] зависимости - функциональные зависимости от ключей.

Зависимость соединения. Пусть A_i - подмножества атрибутов отношения E . Если E эквивалентно соединению проекций $E[A_i]$, то имеет место зависимость по соединению $^*\{A_i\}$. Эта зависимость тривиальная, если одна из проекций $E[A_j]$ совпадает с E .

Определение 5. Отношение находится в 5НФ (проекционно-соединительной НФ), тогда и только тогда, когда каждая его нетривиальная зависимость соединения такова, что каждое подмножество атрибутов этой зависимости является суперключом.
Т.е. $\forall ^*\{A_1, \dots, A_n\}, \forall i=1, \dots, n \exists K_i \subseteq A_i \mid K_i$ - ключ.

Теорема Фейгина (переформулированная). $A \twoheadrightarrow B \mid C \equiv ^*\{AB, AC\}$.

ER-модели

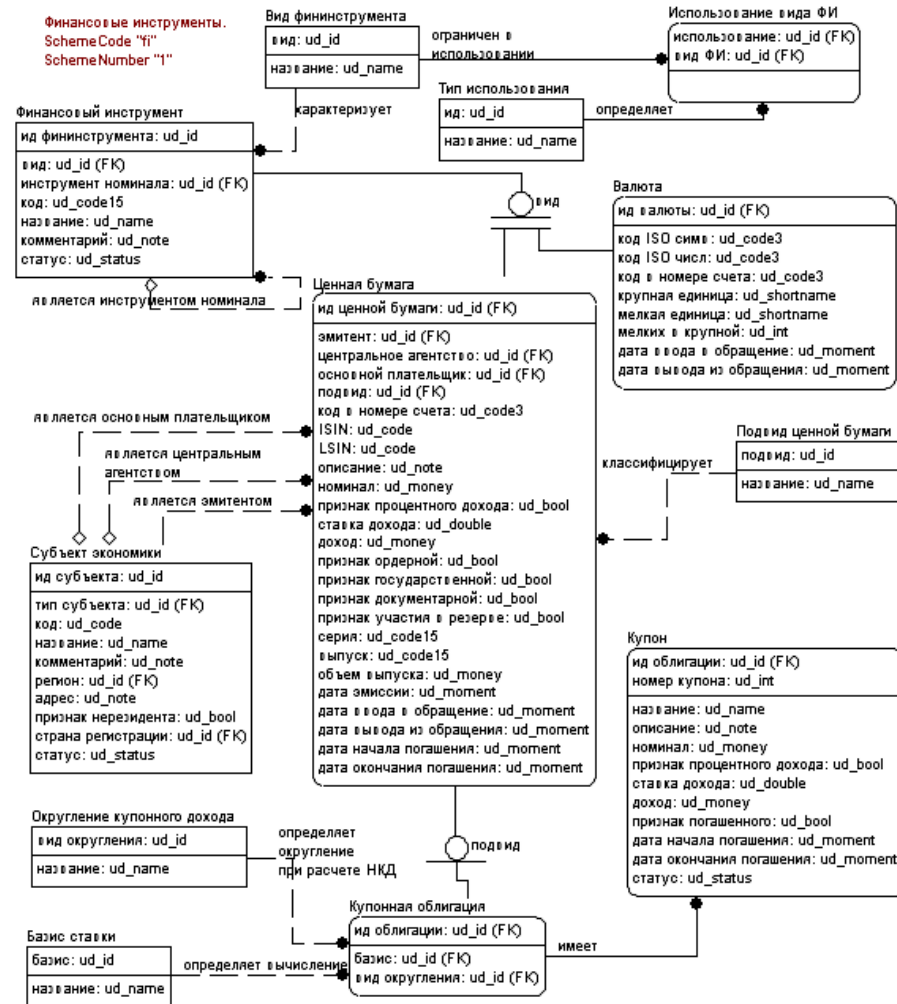
Задача: Представить пул данных в виде набора сущностей (возможно, связанных друг с другом) так, чтобы каждый набор семантически одинаковых атрибутов имел ровно одну сущность-источник.

Пул данных структурирован:

- выделены сущности и их атрибуты;
- выделены функциональные зависимости атрибутов и определены отношения;
- все отношения приведены, как минимум, к 1НФ (т.е. все атрибуты атомарны);
- выделены ключи отношений

ДОПОЛНИТЕЛЬНЫЕ МАТЕРИАЛЫ

Диаграммы IDEF-1x



Пример опросника

Опросник для подготовки предложения по внедрению автоматизированной системы контроля и мониторинга оборотной тары и багажных тележек с использованием технологий радиочастотной идентификации объектов (RFID)

1 ОБЩАЯ ИНФОРМАЦИЯ

- 1.1 Аэропорт:
- 1.2 Код IATA:
- 1.3 Контактная информация:
- 1.4 Пассажиропоток:

2 ОБОРОТНАЯ ТАРА

- 2.1 Типы используемой оборотной тары (багажные тележки, контейнерные тележки и т.д.)
- 2.2 Марки, модели используемой оборотной тары
- 2.3 Количество (отдельно для каждого типа)
- 2.4 Количество оборотной тары, одновременно находящейся:
 - в зоне обработке багажа/грузов
 - в зоне отстоя
 - на перроне
- 2.5 Срок службы оборотной тары (в среднем)
- 2.6 Средняя стоимость единицы оборотной тары

3 ЗОНЫ КОНТРОЛЯ ОБОРОТНОЙ ТАРЫ

- 3.1 Количество зон отстоя оборотной тары
- 3.2 Геометрические размеры зон отстоя
- 3.3 Наличие ограждения в зонах отстоя
- 3.4 Количество въездов/выездов в зоны отстоя (для каждой зоны отдельно указать въезды и выезды)
- 3.5 Количество зон обработки багажа/грузов (загрузка/выгрузка)
- 3.6 Количество въездов/выездов в зоны обработки оборотной тары (для каждой зоны отдельно указать въезды и выезды)

4 БАГАЖНЫЕ ТЕЛЕЖКИ

- 4.1 Желаемые зоны контроля, их расположение и количество (типовые варианты: входы/выходы в залы регистрации, ожидания, переходы между пассажирскими терминалами, выходы на посадку, зоны хранения/отстоя багажных тележек).
- 4.2 Геометрические размеры зон хранения (отстоя) багажных тележек.
- 4.3 Наличие ограждения, фиксированных въездов/выездов в зонах отстоя.
- 4.4 Марки, модели используемых багажных тележек
- 4.5 Стоимость багажной тележки
- 4.6 Количество одновременно использующихся багажных тележек
- 4.7 Средний срок службы багажной тележки
- 4.8 Необходимость средств, препятствующих перемещению багажной тележки за пределы помещений (с указанием количества возможных выездов).
- 4.9 Статистические данные по доступности багажных тележек (при наличии).

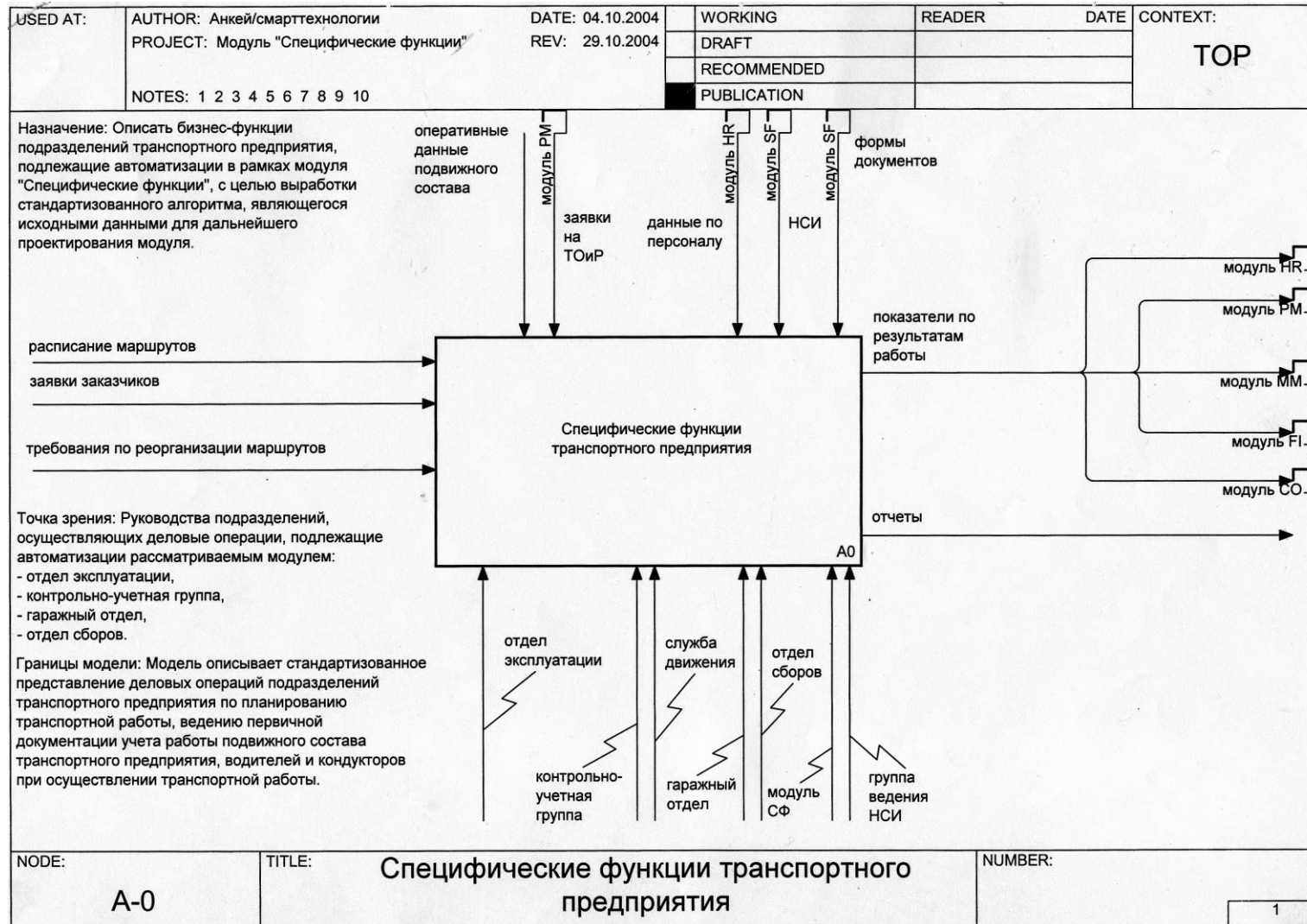
5 КОММУНИКАЦИИ

- 5.1 Наличие электропитания 220 В переменного тока в:
 - в зоне отстоя оборотной тары
 - в зоне обработки багажа/грузов
 - зонах контроля багажных тележек
- 5.2 Наличие ЛВС и возможность развертывания:
 - в зоне отстоя оборотной тары
 - в зоне обработки багажа/грузов
 - зонах контроля багажных тележек
- 5.3 Наличие WiFi и возможность развертывания:
 - в зоне отстоя оборотной тары
 - в зоне обработки багажа/грузов
 - зонах контроля багажных тележек

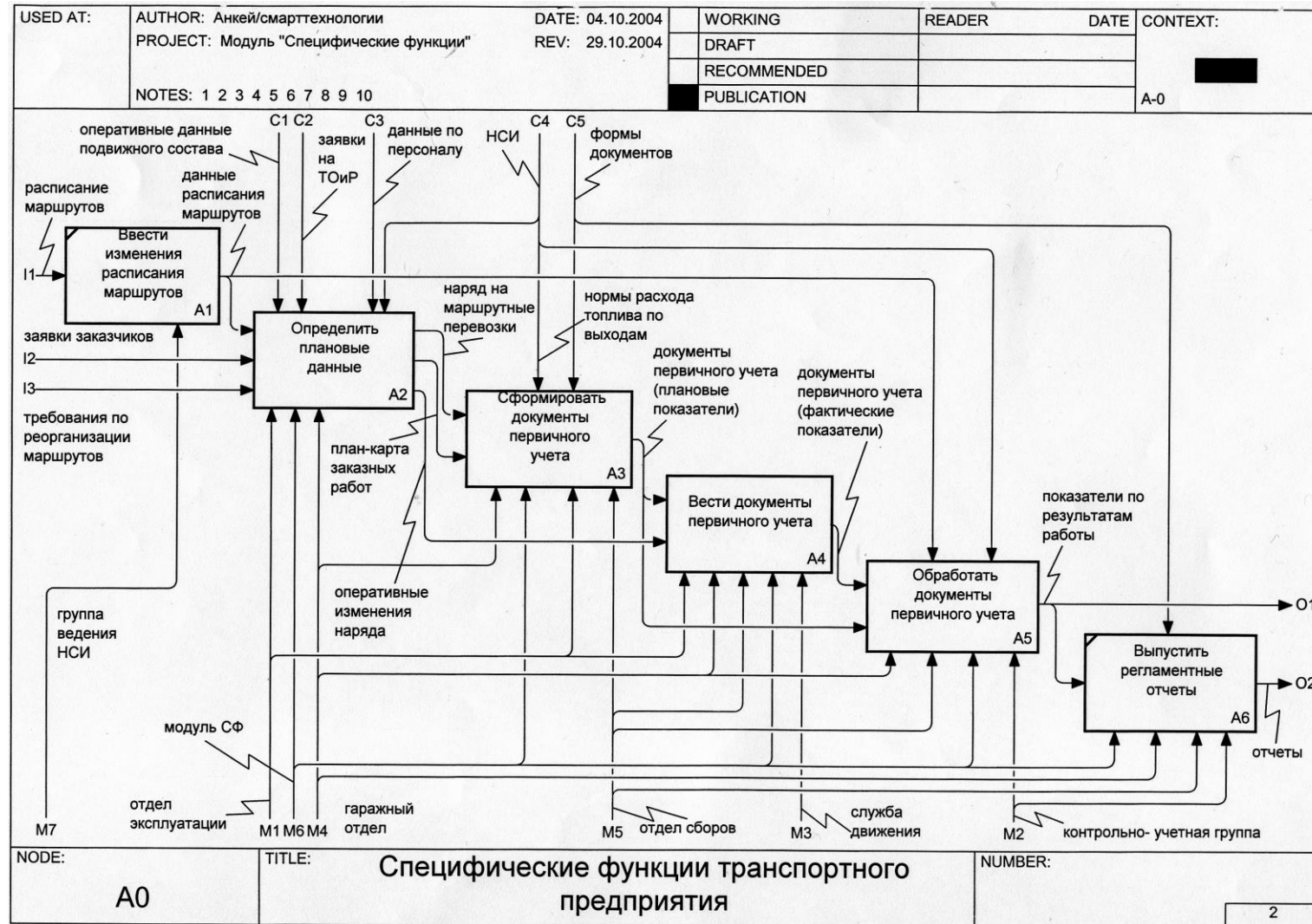
6 ИНФОРМАЦИОННОЕ ВЗАИМОДЕЙСТВИЕ С ВНЕШНИМИ СИСТЕМАМИ

- 6.1 Необходимость информационного обмена RFID-систем с ERP системой предприятия (например, SAP).
- 6.2 Используются ли на предприятии модули SAP APO, SAP TM или их аналоги.

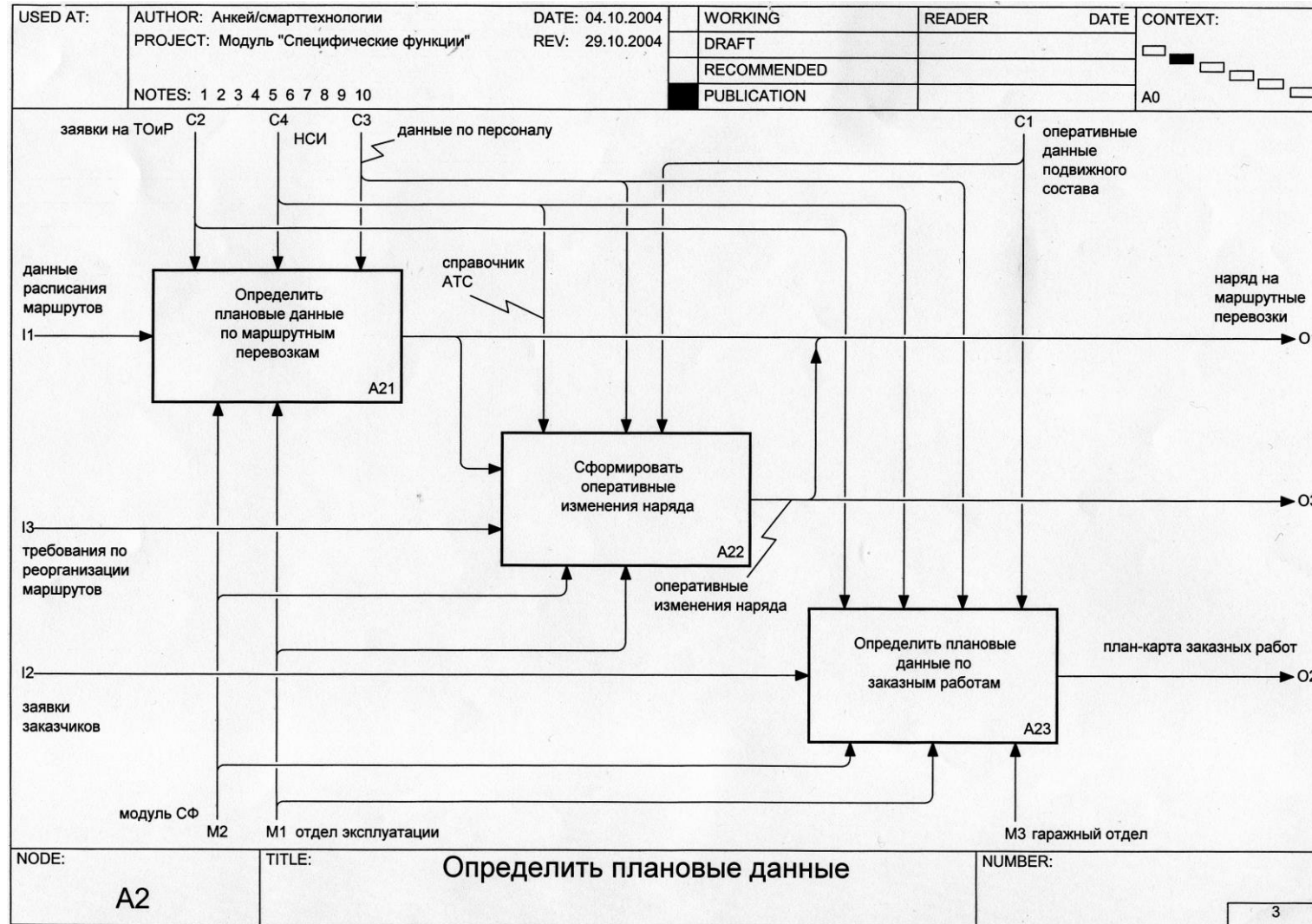
Диаграммы IDEF-0



Диаграммы IDEF-0



Диаграммы IDEF-0



Диаграммы IDEF-0

