

Общая модель Системы открытого архива (OAIS)

Сохранность Цифровых объектов – принципы и практика

DPE, Planets CASPAR и Nestor

Prague, Czech Republic, October 13-17, 2008

Carlo Meghini

Consiglio Nazionale delle Ricerche

Istituto di Scienza e Tecnologie della Informazione

<http://nmis.isti.cnr.it/meghini/>

Acknowledgements

Michael Day

Digital Curation Centre

UKOLN, University of Bath

<http://www.ukoln.ac.uk/>

Session outline

- Background
- Mandatory Responsibilities
- Functional Model (repository view)
- Information Model (object view)

OAIS background

- Справочная модель Open Archival Information System (OAIS)
- Разработка осуществлена консультативным комитетом космических информационных систем (CCSDS)
- Issued as CCSDS Recommendation (Blue Book) 650.0-B-1 (January 2002)
- Also adopted as: ISO 14721:2003
- Periodic reviews
- <http://public.ccsds.org/publications/archive/650x0b1.pdf>

Задачи и диапазон применения OAIS (1)

- Определение системы открытого архива (OAIS)
 - An OAIS - сложная система, обеспечивающая хранение информации и ее доступность для определенного сообщества.
 - Слово 'открытый' в названии означает, что принципы разработки системы открыты. Однако, оно не подразумевает неограниченного доступа к любой системе OAIS
 - Сама по себе OAIS не является неизменяемой системой, но хранящаяся в ней информация рассматривается как информация, подлежащая «длительному хранению»
 - Длительное хранение = хранение продолжительное настолько, чтобы возникла необходимость учитывать влияние меняющихся технологий.

Задачи и диапазон применения OAIS (2)

- Исходная ориентация на цифровую информацию
 - Хранению подлежат как первичные формы информации, так и поддерживающая информация как для цифровых так и для физических архивов.
- Модель содержит информацию, которая изначально нецифровая (напр., физические единицы хранения)
 - Но модели и сохранность такой информации не рассматриваются в деталях.

Задачи и диапазон применения OAIS (3)

- Специальные цели:
 - Основная концепция – что подлежит длительному хранению и доступу
 - Терминология и концепты для описания и сопоставления:
 - Структура и операции
 - Стратегия и технология обеспечения сохранности
 - Модели данных
 - Соглашения об элементах и процессах по обеспечению долговременного хранения и доступа, а также расширение рынка
 - Основы других стандартов
 - Информация НЕ в цифровой форме
 - OAIS- связи

Задачи и диапазон применения OAIS (4)

– Применение:

- Применима к любому архиву, но в основном в организациях, отвечающих за создание информации доступной в течение длительного времени
- Представляет интерес для тех, кто создает информацию, которая может нуждаться в длительном хранении и для тех, кто приобретает информацию из таких архивов.
- В данном случае не указывается конкретный проект или тип внедрения. В каждом случае функции могут группироваться или разделяться.

– Указаны соответствующие стандарты (секция 1.5)

Задачи и диапазон применения OAIS (5)

– Требования

- OAIS должна поддерживать информационную модель
- Обязательные требования (раздел 3.1)
- Сама по себе модель технологически независима
 - «Предполагается, что разработчики используют данную модель в качестве руководства в ходе разработки конкретной системы с определенными сервисами и контентом»
 - Модель не предполагает специфической программно-аппаратной платформы, системной среды, проектной парадигмы, СУБД, языка описания данных и т.д..
 - OAIS может поддерживать дополнительные сервисы

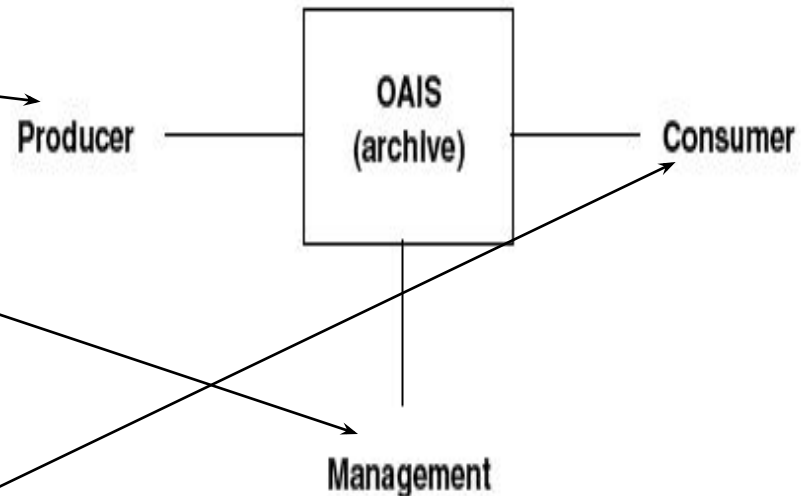
OAIS концепты высшего уровня (1)

- Под традиционными архивами понимают системы или организации, которые сохраняют материалы с целью предоставления к ним доступа.
 - Архивы осуществляют это назначение, убедившись, что подлежащие хранению материалы представляют интерес для сообщества пользователей, которому они предназначены, и взяв соответственно, на себя ответственность за сохранность их информационного содержания и аутентичность, а также осуществляя управление ими и предоставляя к ним доступ.
- Поскольку цифровая информация легко утрачивается или портится, обеспечение ее сохранности требует совместных усилий большого государственных, коммерческих и некоммерческих организаций

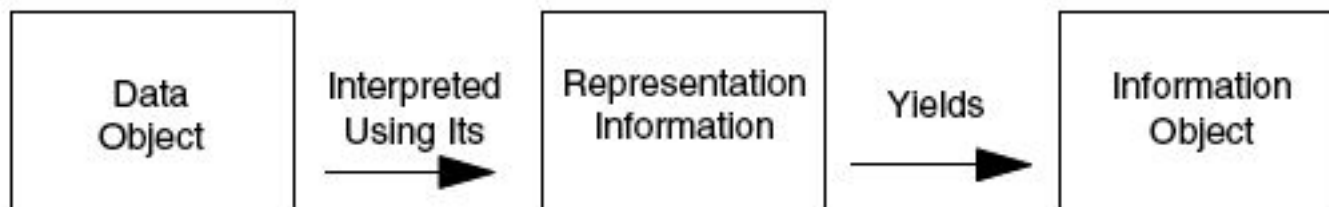
OAIS концепты высшего уровня (2)

- OAIS окружение

- Производитель предоставляет информацию
- Менеджмент определяет общую стратегию (не каждодневные операции)
- Пользователь (потребитель) находит и приобретает информацию, представляющую интерес
 - Выделенная группа пользователей – это сообщество пользователей, способных понять сохраненную информацию



- Можно сказать, что лицо или система обладает базой знаний, позволяющей понять получаемую информацию.
- Информация – знание любого типа, которым можно обмениваться, и которое выражено в некотором типе данных.
 - Информация в книге обычно выражена знаками (данными) которые, в сочетании со знанием используемого языка (База знаний), преобразуются в значимую информацию. Если реципиенту язык не знаком, книга должна сопровождаться словарем и грамматикой (т.е, **Информацией о представлении**) в форме понятной с помощью базы знаний реципиента



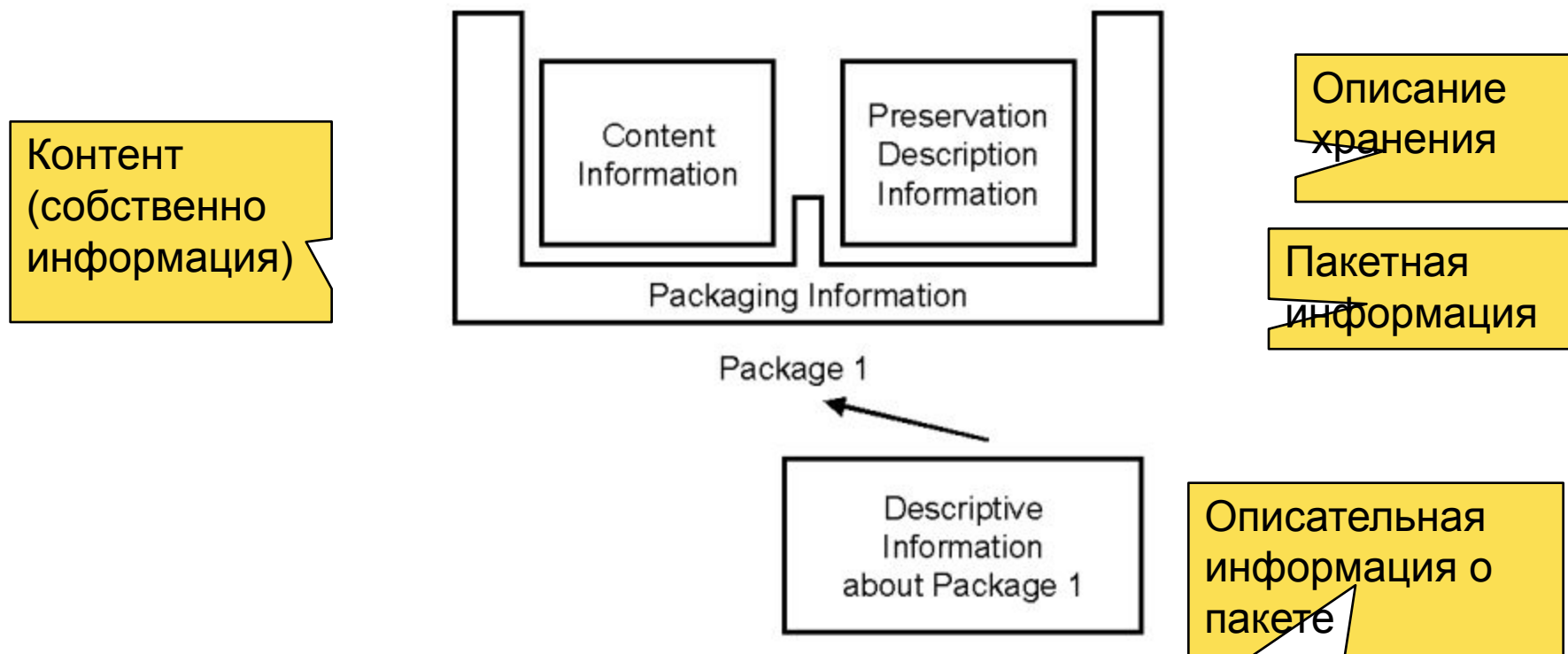
OAIS концепты высшего уровня (4)

- Для того, чтобы данный информационный объект был успешно сохранен, необходимо четко представлять себе объект данных и связанную с ним информацию о представлении.
- Необходимо также представлять себе базу знаний выделенной группы пользователей, чтобы представлять себе минимум информации представления, который должен поддерживаться.

OAIS концепты высшего уровня (5)

- Единицей обмена между OAIS и ее окружением является информационный пакет.
- Информационный пакет – это концептуальный контейнер двух типов информации:
 - Контент и
 - Описание хранения - Preservation Description Information (PDI).
- Общий пакет должен открываться с помощью описательной информации

OAIS концепты высшего уровня (6)



Information Package Concepts and Relationships (Figure 2-3)

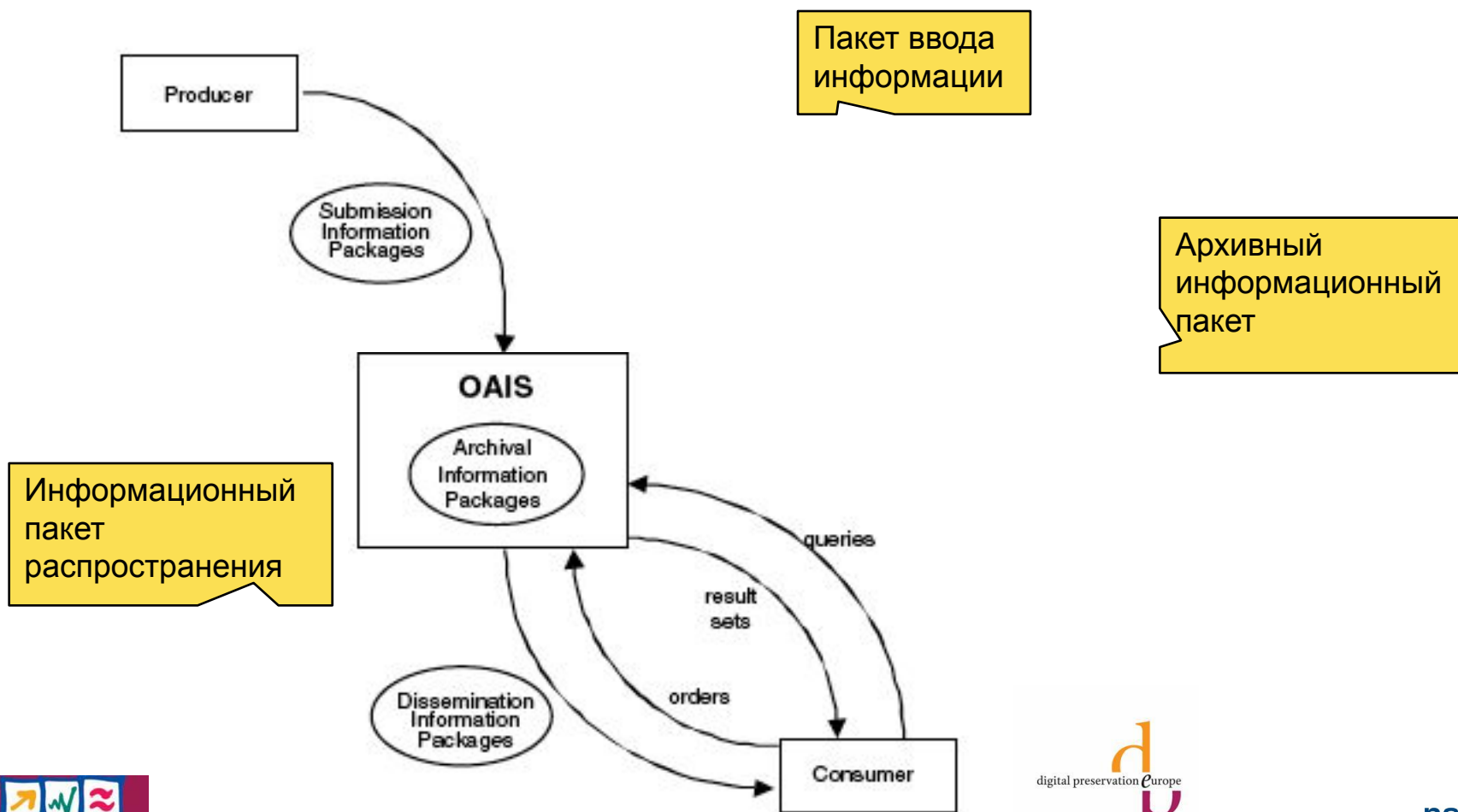
OAIS концепты высшего уровня (7)

- Пакетная информация (Packaging Information) это та информация которая, описательно или логически, устанавливает, идентифицирует и связывает Контент and PDI (Preservation Descriptive Information).
- Описательная информация (Descriptive Information) используется для выявления пакета, контент которого представляет интерес.

OAIS концепты высшего уровня(8)

- Типы информационного пакета
 - Пакет ввода информации (Submission Information Package (**SIP**))
 - Архивный информационный пакет (Archival Information Package (**AIP**))
 - Информационный пакет распространения (Dissemination Information Package (**DIP**))
- Пакеты могут варьировать в зависимости от их роли
 - Например , проекты графических изображений и электронных журналов часто различаются по содержанию хорошо управляемых (и описанных) «мастер» файлов и их производных (иконок, JPEG , PDF), доступных через Интернет

OAIS внешние взаимодействия (1)



OAIS внешние взаимодействия (2)

- Взаимодействия высокого уровня в окружении OAIS
 - Взаимодействие менеджмента
 - Финансирование, оценка, решение конфликтов
 - Взаимодействие производителя
 - Соглашение о вводе
 - Взаимодействие пользователя
 - Справочная служба, выявление информации, заказ информации

OAIS обязанности и ответственность (1)

- Организовывать поставку и принимать информацию от производителей
- Иметь достаточную систему управления информацией на уровне, необходимом для обеспечения длительного хранения
- Определить, самостоятельно или совместно с другими коллегами, для какой группы пользователей предназначена информация

OAIS обязанности и ответственность (2)

- Сохраняемая информация должна быть понятна выделенной группе пользователей

Пользователи должны осваивать информацию самостоятельно без помощи каких либо экспертов, производителей информации или каких-либо других лиц

- Необходимо следовать документированной стратегии и процедурам, которые:

- а) обеспечивают сохранение информации в любых прогнозируемых условиях, и
- б) создают возможность распространения информации в виде аутентичных копий оригинала, или в виде ссылок к оригиналу

OAIS обязанности и ответственность (3)

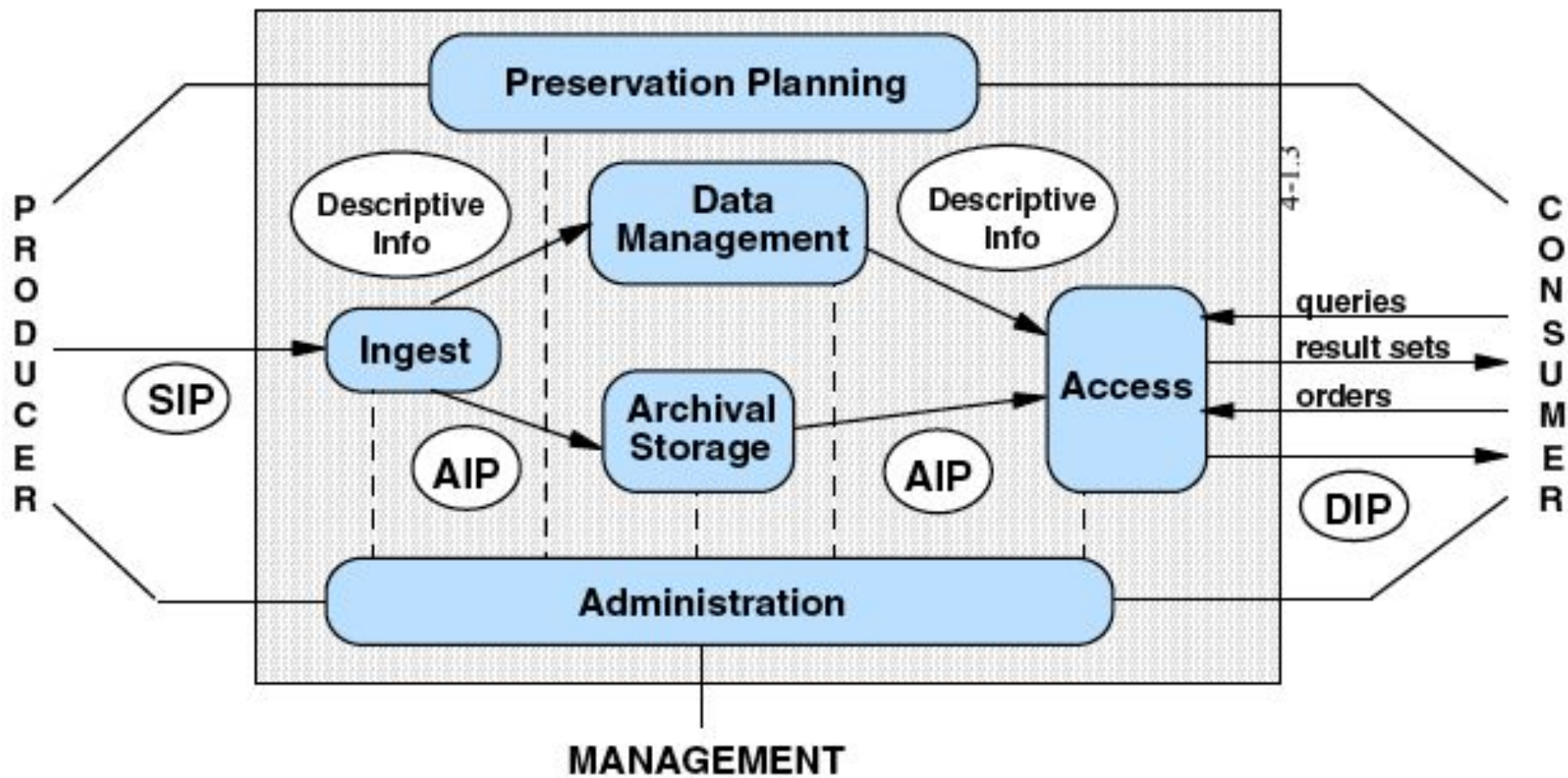
- Обеспечить доступность информации выделенной группе пользователей.
- Секция 3.2 описывает механизмы выполнения обязанностей

Функциональная модель OAIS

(раздел 4.1)

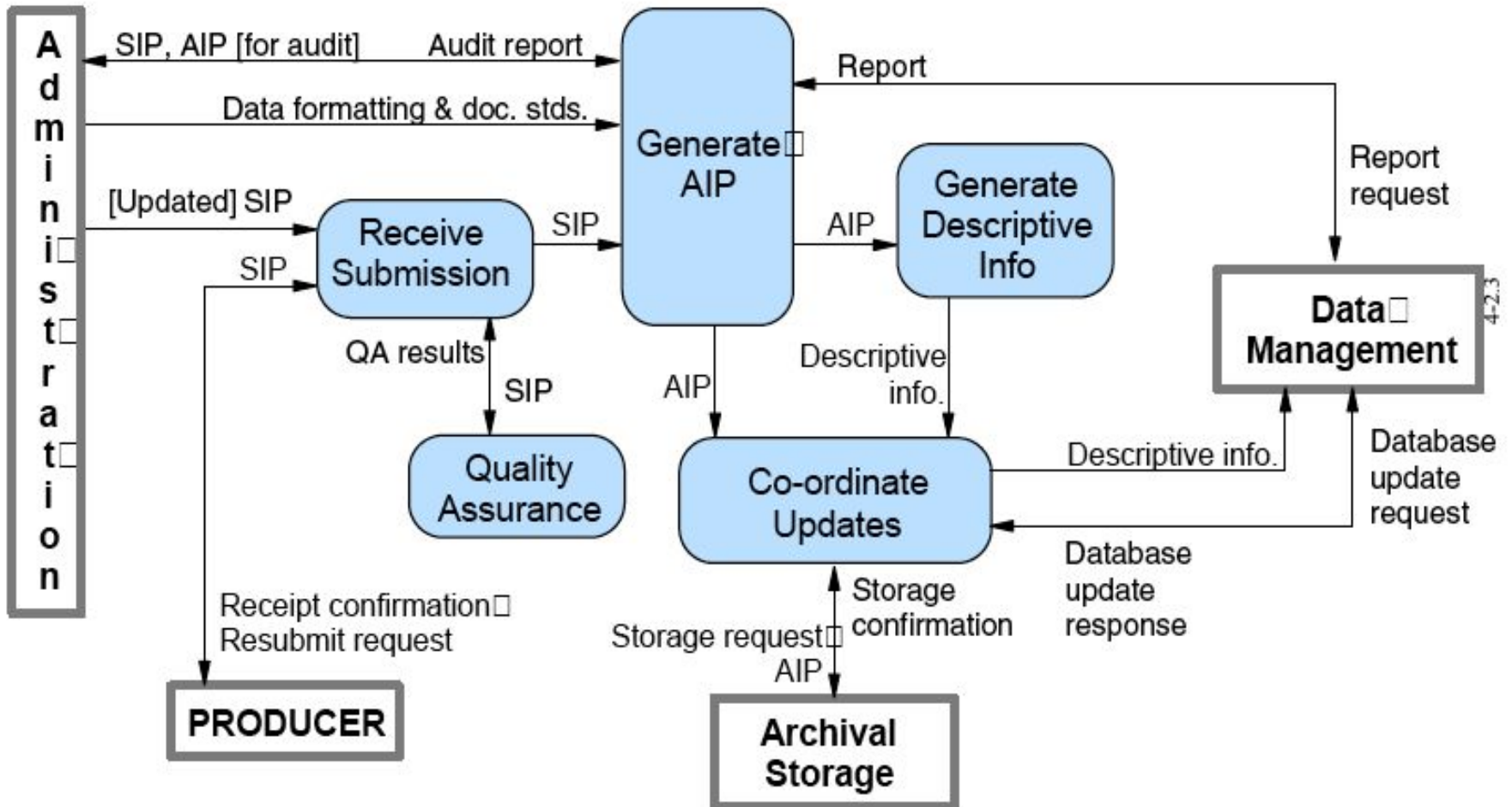
Функциональная модель OAIS (1)

- Шесть функциональных модулей и соответствующих интерфейсов
 - Ввод
 - Архивное хранение
 - Управление данными
 - Администрация
 - Планирование хранения
 - Доступ
- Описание, диаграммы...



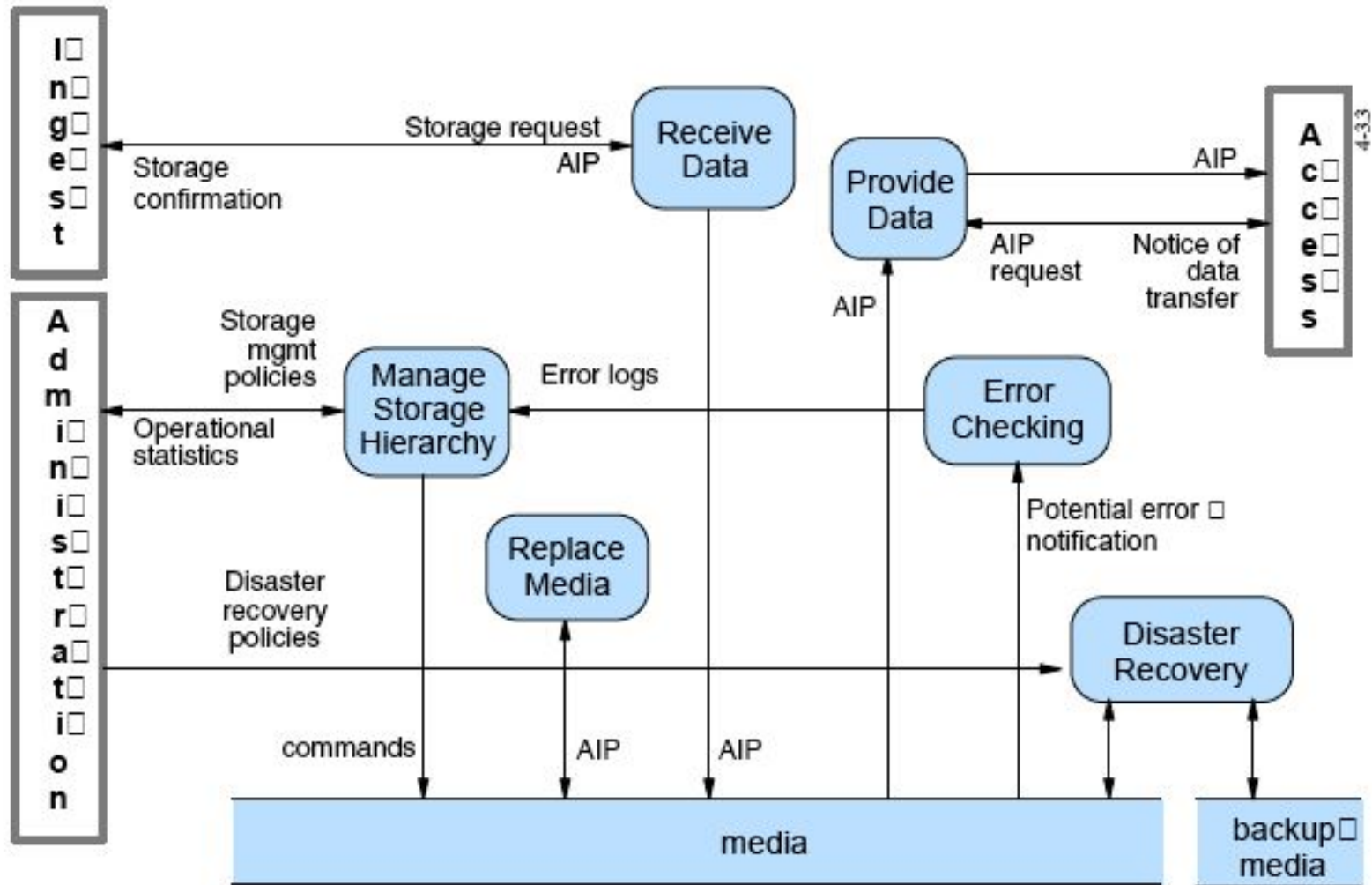
Ввод (Ingest)

Обеспечивает сервисы и функции Information Пакета ввода информации (SIPs) от создателей (или из внутренних элементов системы под контролем администрации) и подготовку контента для хранения и управления в архиве.



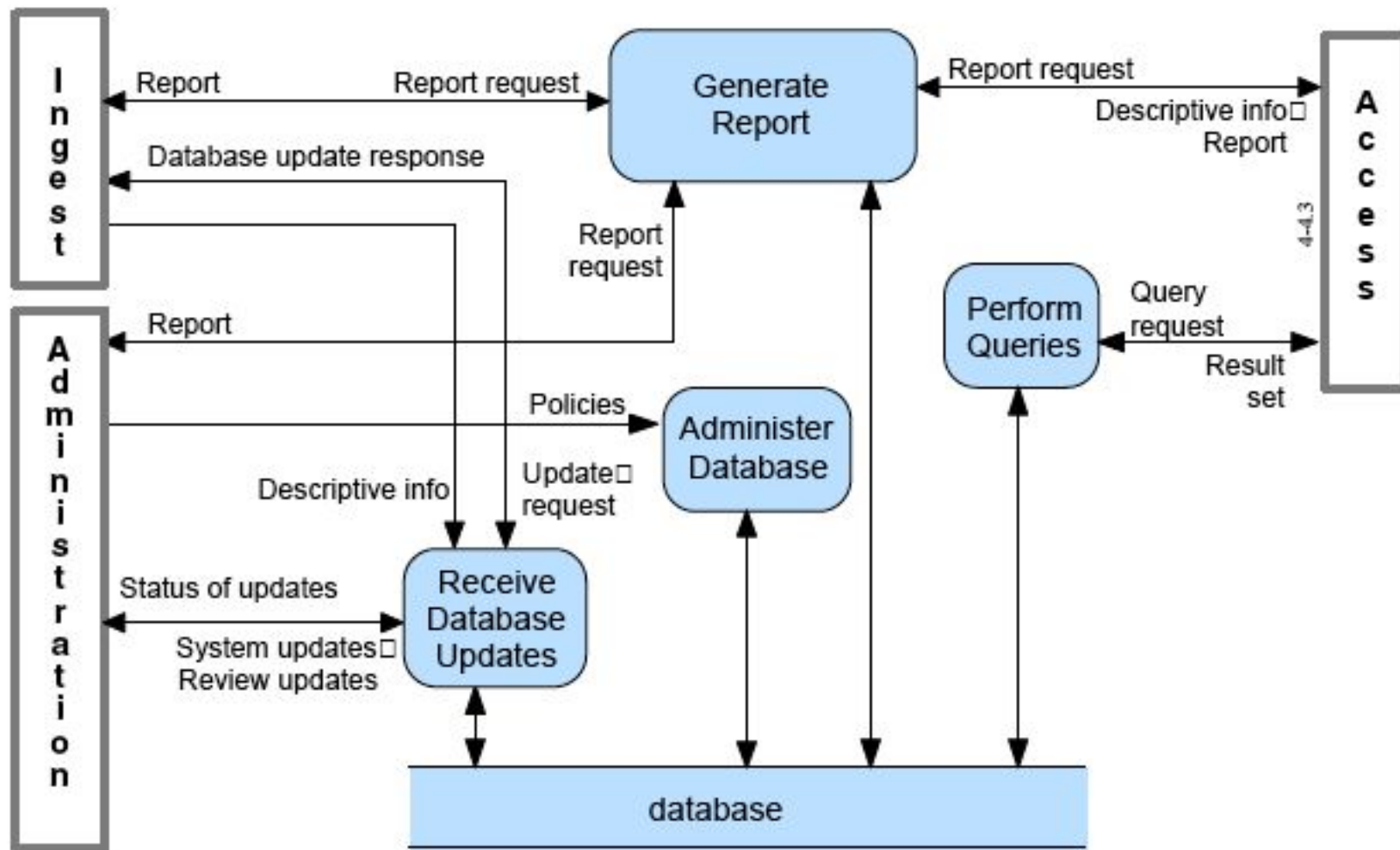
Архивное хранение

Обеспечивает сервисы и функции хранения, поддержки и извлечения архивных пакетов.



Управление данными (Data Management)

Обеспечивает сервисы и функции для поддержки и доступа как описательной информации которая идентифицирует данные и документы архива и к административным данным .



Администрирование (Administration)

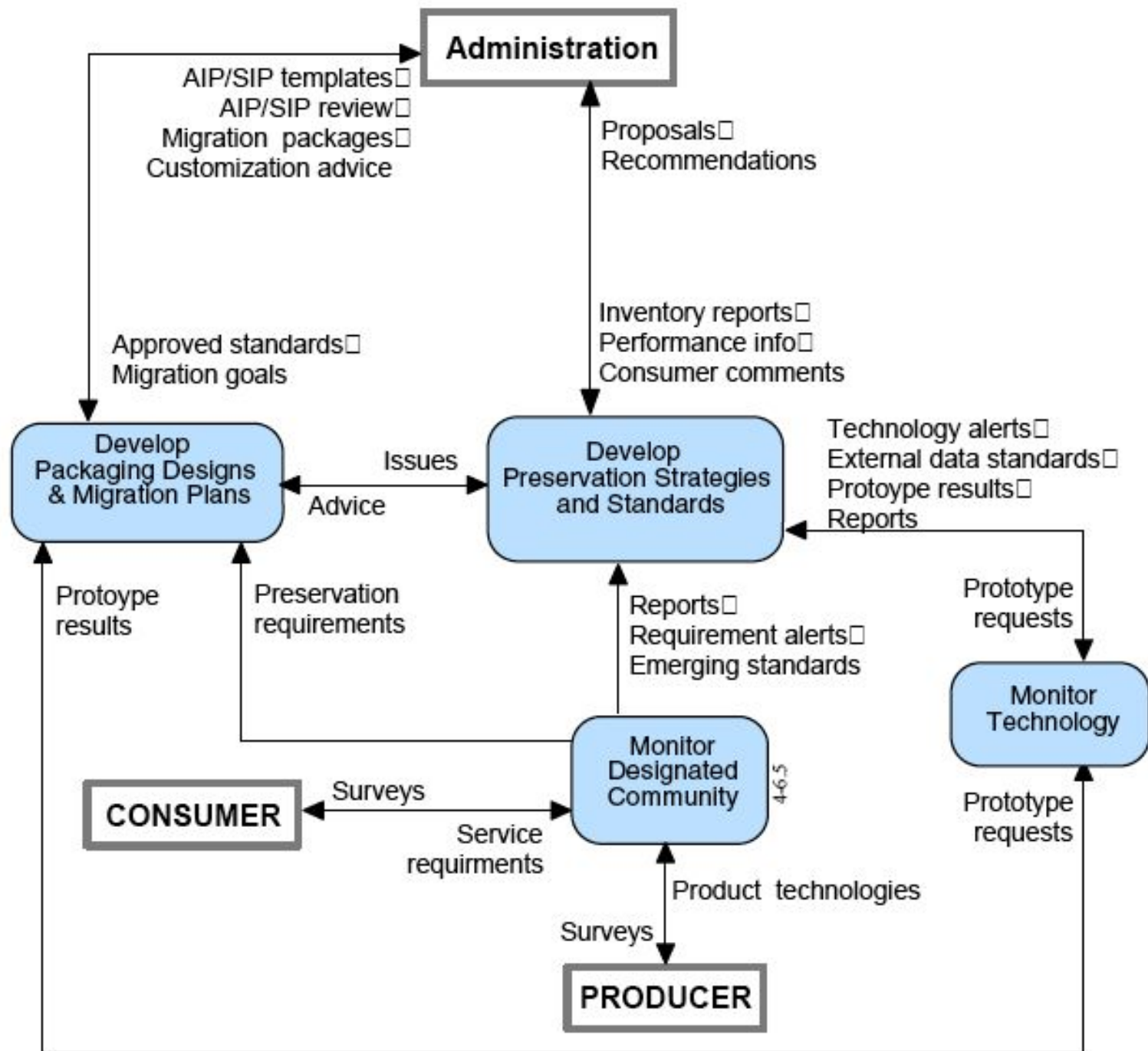
Обеспечивает общие операции архивной системы, в т.ч.

:

- Организационная работа по заключению соглашений
- Проверка ввода, на предмет соответствия стандартам
- Поддержка конфигурации системы.

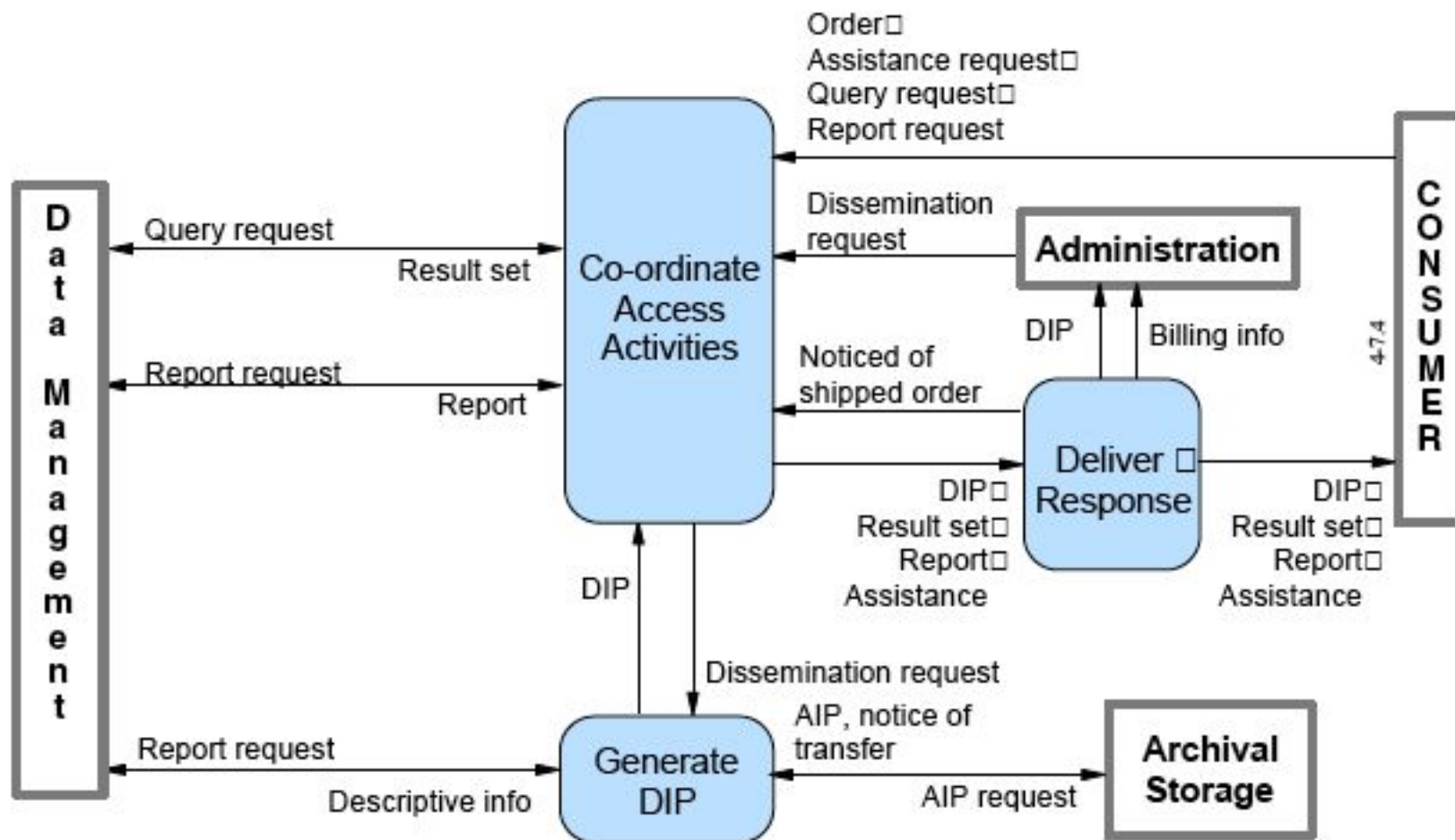
Планирование хранения

Обеспечивает сервисы и функции для мониторинга окружения OAIIS и обеспечения рекомендаций с целью того, чтобы информация хранящаяся в OAIIS оставалась доступной обозначенному сообществу пользователей в течение длительного времени, даже в том случае, если оригинальная компьютерная среда устареет.



Доступ (Access)

Обеспечивает сервисы и функции, поддерживающие пользователей при определении наличия, описания, местонахождения и доступности информации хранящейся в OAIS, и позволяющие пользователям запрашивать и получать информационные продукты.



Информационная модель OAIS

(Section 4.2)

Введение

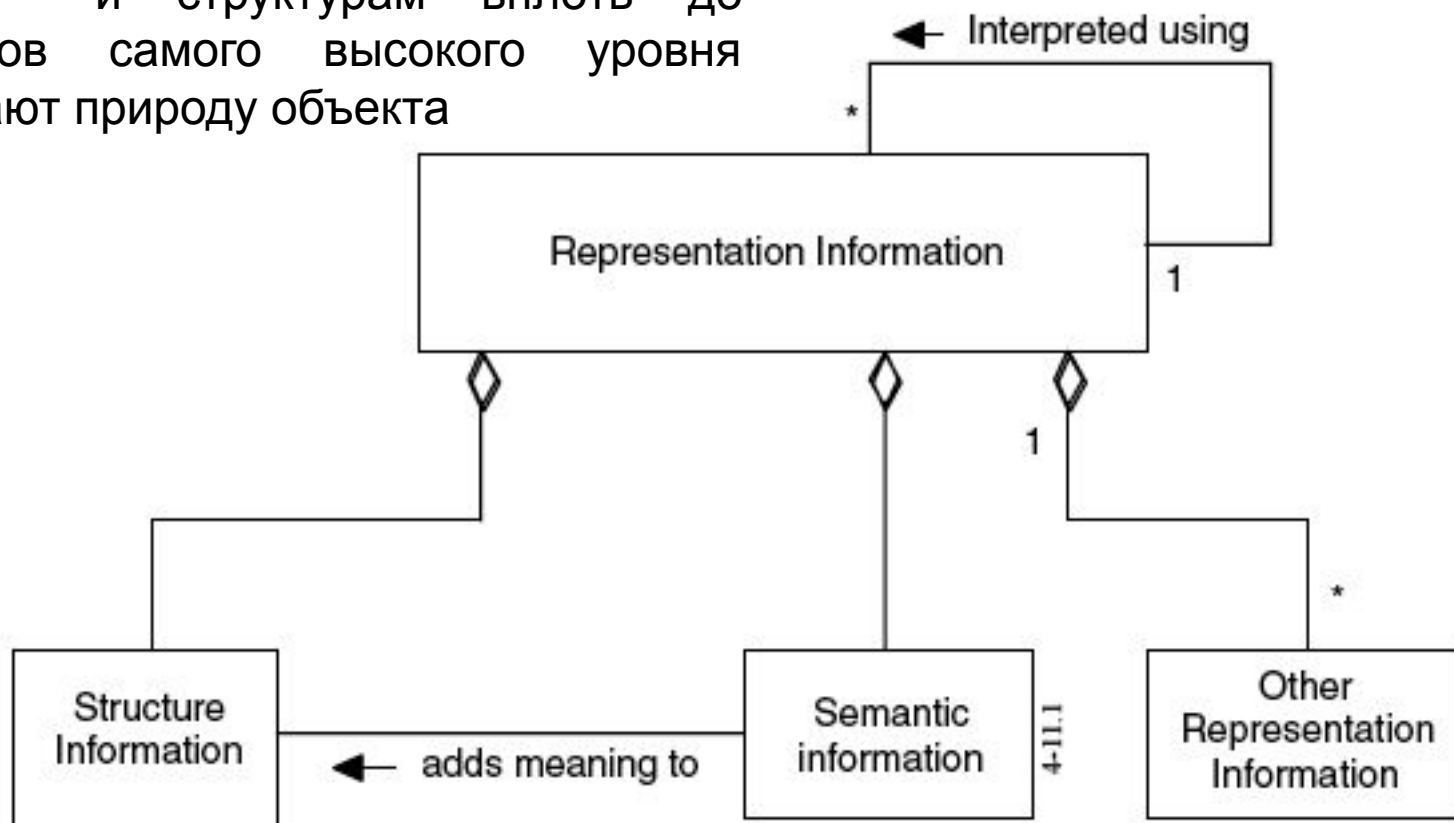
- Основная цель OAIS сохранить информацию для выделенной группы пользователей на неопределенный период времени.
- В этом смысле OAIS должен сохранить значительно больше, чем просто контент объектов, которые подлежат хранению.
- Информационная модель описывает типы информации, которая хранится в OAIS .



Представление информации может дать ей дополнительное значение посредством

- (1) Обеспечивая соответствие битов распознаваемым типам данных (буквы, цифры, строки, записи, и т. д.);
- (2) устанавливая связи со значениями более высокого уровня, которые получают определения и взаимосвязи в онтологиях.

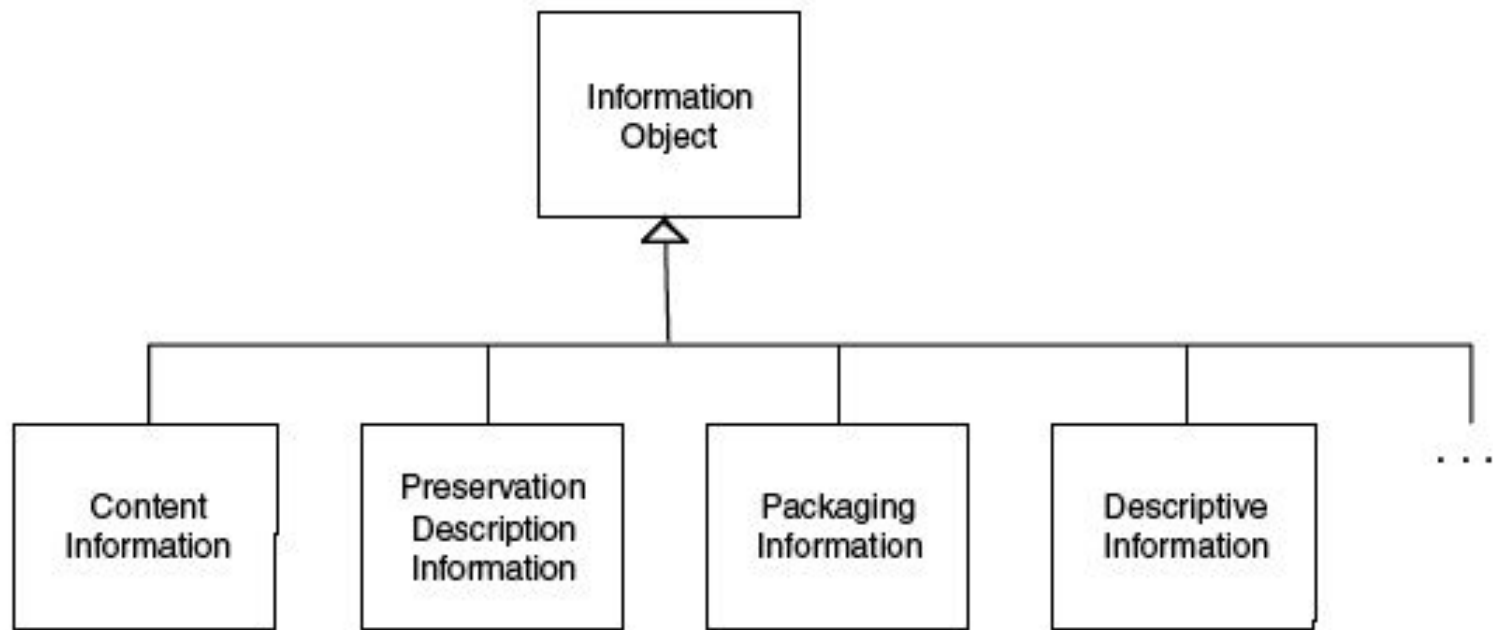
Правила соответствия битов значениям данных и структурам вплоть до концептов самого высокого уровня учитывают природу объекта



Представление. Информация. Сети

- Представление информации может содержать ссылки к другим представлениям информации
- Представление информации само по себе является информационным объектом, который также может иметь свой цифровой объект и другое представление информации для понимания цифрового объекта
- Результирующее множество объектов можно назвать сетью представления.

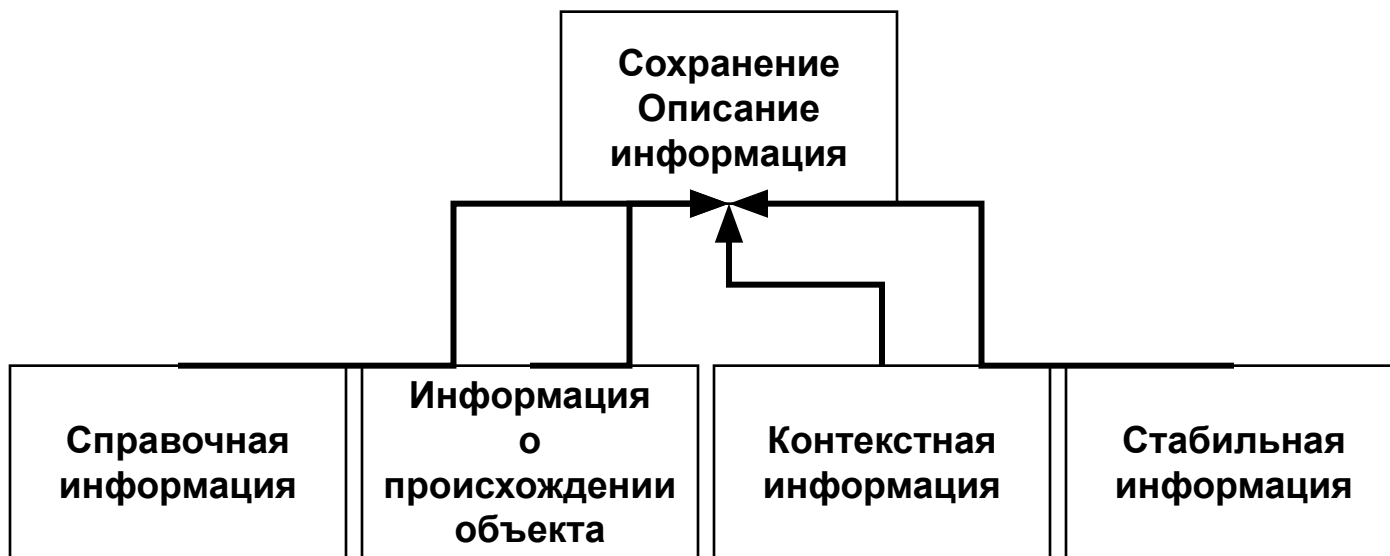
Типы информационных объектов



Контент

- Контент – это информационный массив (информационное наполнение), который является изначально объектом хранения OAIS.
- Контент – информация это объекты данных вместе с информацией об их представлении. Таким объектом может быть цифровой или физический объект.
- Любой информационный объект может являться контентом.

Сохранение Описание Информация



PDI Preservation Description Information (Figure 4-16)

Сохранение Описание Информация

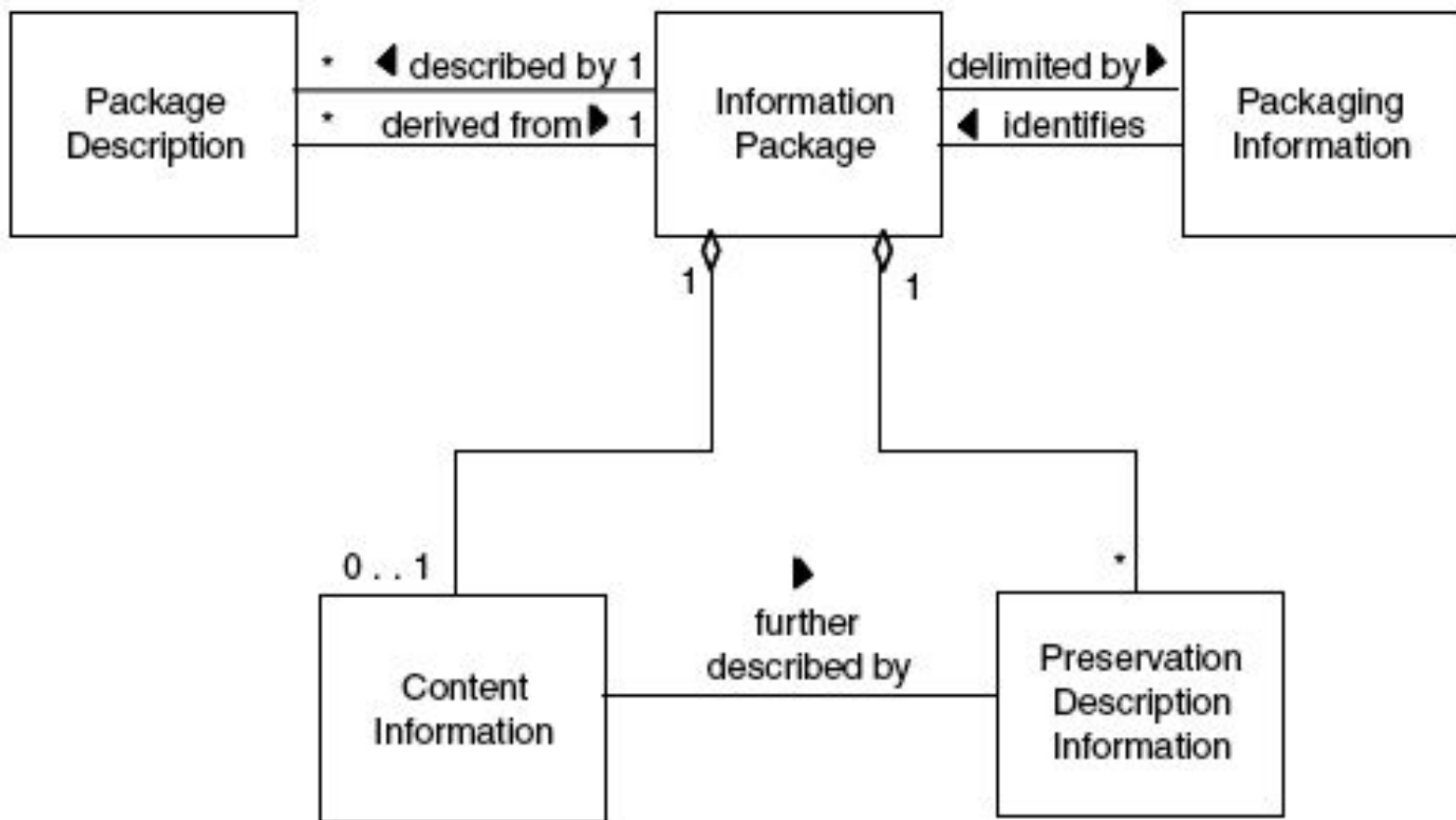
- **Ссылочная информация:** идентифицирует и описывает один или более механизмов, используемых для присвоения контенту идентификаторов. Она также описывает сами идентификаторы.
- **Контекстная информация:** документы, содержательно связанные с контентом, или устанавливающая связи контента со средой (зачем создан данный контент, и как он связан с другим контентом).

Сохранение Описание Информация

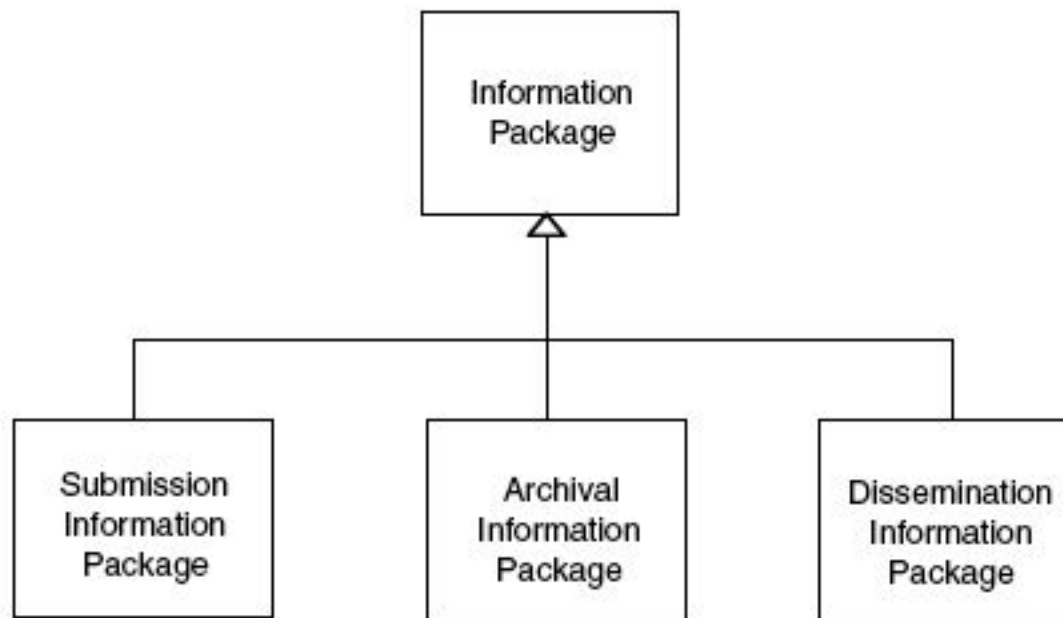
- **Информация о провенансе:** документирует историю контента (происхождение или источник, изменения и места хранения). Провенанс может рассматриваться как особый тип контекстной информации.
- **Информация о стабильности:** описывает целостность данных или критерии определения того, что данный информационный объект не был изменен без соответствующего документирования.

Информационные пакеты OAIS

- Концептуальные информационные структуры необходимы для выполнения функций OAIS.
- Информационный пакет представляет собой контейнер.
- Существуют несколько типов информационных пакетов, которые используются в ходе работы архива.
 - Для структурирования и хранения материалов OAIS (AIP);
 - Транспортировка от изготовителя в архив (SIP)
 - Транспортировка из архива к пользователю (DIP).

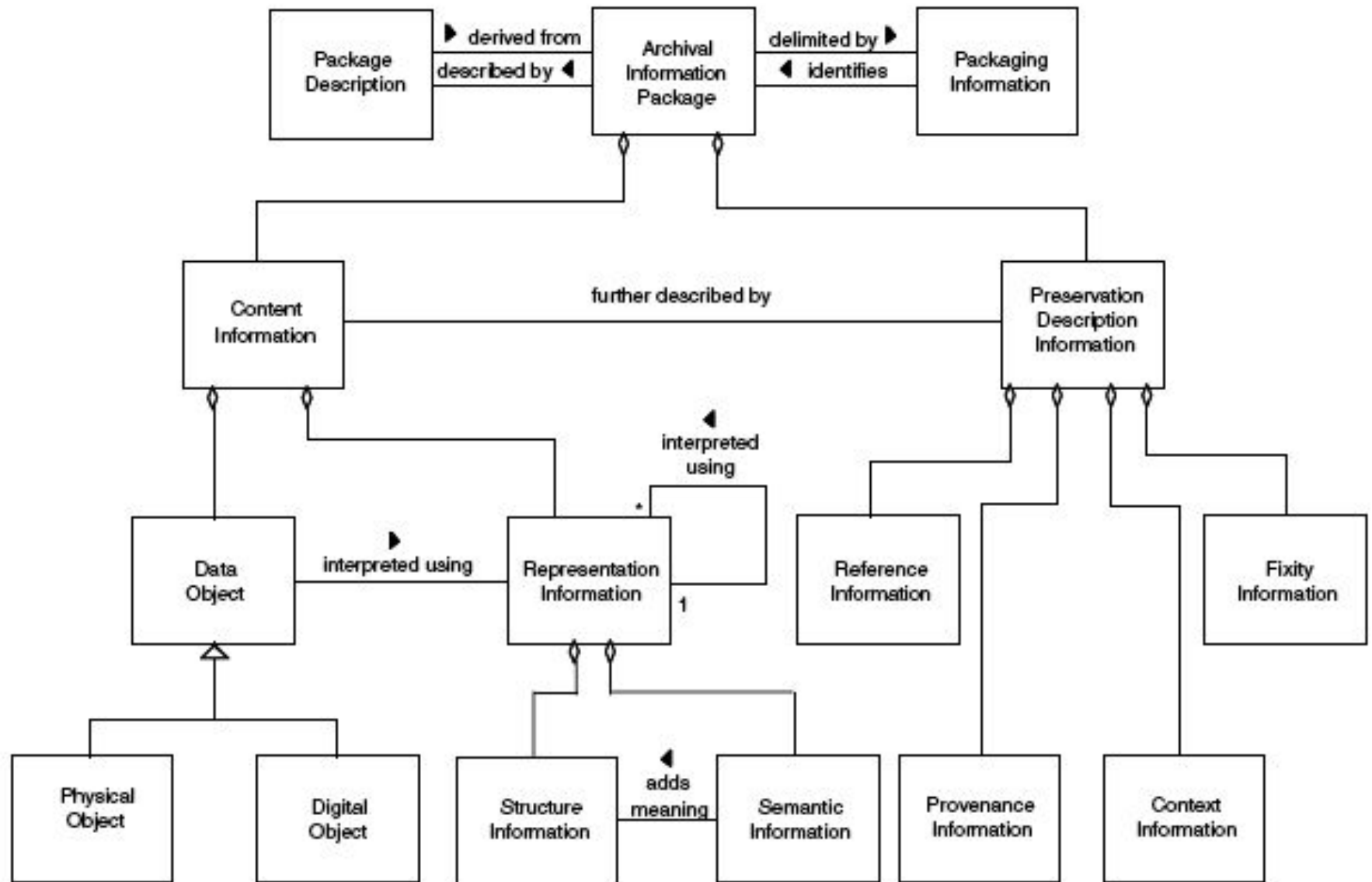


Типы информационных пакетов



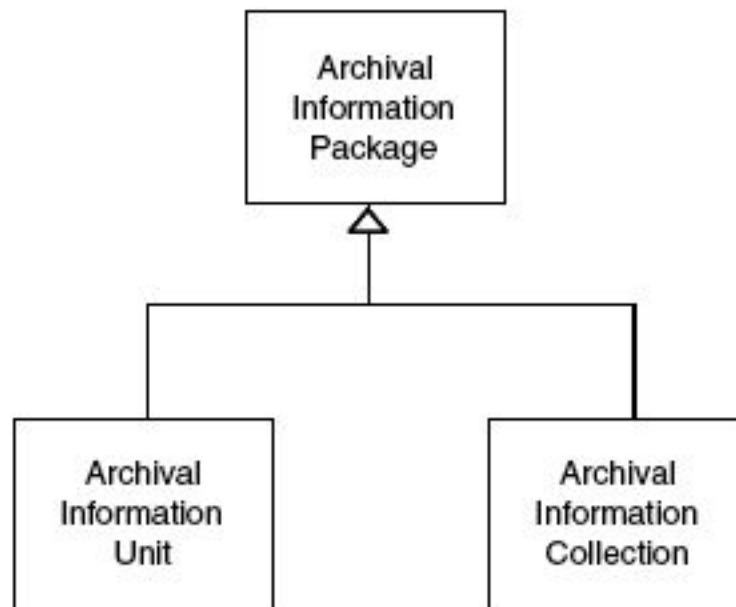
Информационный пакет ввода (SIP)

- Форма и детальное содержание пакета ввода (SIP) обычно согласовываются производителем и архивом OAIS.
- Большинство пакетов ввода содержат некоторый контент и некоторое описание хранения (PDI). Для передачи всего контента и связанных с ним описаний могут понадобиться несколько пакетов ввода.
- Если имеются множественные пакеты ввода, которые используют один и тот же RepInfo, скорее всего этот RepInfo быть представлен один раз.
- В пределах OAIS, пакеты ввода (SIPs) числом один или более трансформируются в архивные пакеты хранения, числом один или более.



Типы архивных пакетов

Под единицей архивной информации (AIU) рассматривается единственный информационный объект, описанный одним набором информации по хранению (PDI).



Собрание архивной информации рассматривается как собрание других собраний и единиц, каждое (или каждая) из которых имеет собственное описание хранения (PDI). Кроме того, само собрание имеет свое описание, содержащее критерии включения и правила.

Пакет распределения информации (DIP)

- В ответ на заказ пользователя, OAIS передает все части архивного пакета потребителю в форме распределительного пакета (DIP).
- Пакет распределения (DIP) может также включать собрания архивных пакетов, в зависимости от соглашения между OAIS и потребителем.
- Пакетная информация должна всегда быть представлена так чтобы потребитель мог ясно видеть заказанную им информацию.
- Назначение описательной информации в пакете – дать потребителю достаточно полное представление об отличии данного пакета от других сходных пакетов.

OAIS – другие перспективы

– Хранение

- Миграция, напр. refreshment (освежение), replication (дублирование), repackaging (переупаковка), трансформация (transformation)
- Внешняя сохранность (напр., эмуляция, виртуальные машины)

– Взаимодействие архивов

- Взаимодействие OAIS архивов (например, кооперирование и федерации архивов)

Внедрение модели OAIS

Основы внедрения (1)

- OAIIS общая модель (концептуальные рамки), НЕ чертеж-проект системы
- Он дает информацию об архитектуре системы, разработке систем и ее компонентов
- Терминология предметной области
- Но он обеспечивает устойчивости и взаимодействия между системами

Основы внедрения (2)

- ISO 14721:2003
 - Следует рекомендациям сделанных доступными CCSDS
 - Однако, существуют более ранние версии модели
- Области применения:
 - Связанные стандарты (e.g., CCSDS Archive-Producer Interface)
 - Стандартизация терминологии
 - Сертификация и следование стандартам
 - Анализ сопоставление архивов
 - Дизайн системы
 - Сохранность метаданных

Соответствие и сертификация

Соответствие OAIS (1)

- Многие хранилища или устройства для сохранения претендуют на соответствие или влияние OAIS :
 - e.g., IBM DIAS, DSpace, OCLC Digital Archive, METS, список бесконечен
 - Система LOCKSS выпустила «формальное заявление о соответствии ISO 14721:2003» (lockss.stanford.edu/)
- Собственное видение модели OAIS (OAIS 1.4):
 - Поддержка информационной модели (OAIS 2.2),
 - Выполнение шести обязательных функций (OAIS 3.1)

Соответствие OAIS (2)

- OAIS обязательные функции:
 - Приобретение и ввод информации
 - Приобретение достаточного контроля информации подлежащей длительному хранению
 - Определение «целевой группы пользователей»
 - Обеспечение независимого понимания информации
 - Следование стратегии и процедурам
 - Обеспечение доступности сохраняемой информации

Сертифицированный (Trusted) цифровой архив (1)

- OCLC/RLG Рабочая группа по цифровому архиву
 - Доклады Trusted Digital Repositories report
<http://www.oclc.org/research/activities/past/rlg/trustedrep/repositories.pdf>
 - Рекомендованный порядок сертификации цифрового архива
 - Модель аудита
 - Модель стандарта
 - Построение на основе модели OAIS ...

Сертифицированный (Trusted) цифровой архив (2)

– Специальные атрибуты:

- Соответствие OAIS
- Административная ответственность
- Организационная устойчивость
- Финансовая надежность
- Технологическое и процедурное соответствие
- Защищенность системы
- Процедурная контролируемость

Сертификация хранилища цифровых материалов

- Тростовый аудит и сертификация хранилищ (TRAC): Критерии и контрольный список (Март 2007)

Организационная инфраструктура

- напр., управление, организационные структуры, стратегические концепции, системы финансирования, контракты и лицензии

Управление цифровыми объектами (OAIS functions)

- напр., ввод, метаданные, стратегии хранения
- Технологии, Технологические структуры & безопасность

Анализ и сопоставление хранилищ

Анализ существующих сервисов

Сопоставление существующих процессов и сервисов с функциональной и информационной моделями OAIS

Основные цели:

- Выявление существенных дефектов системы
- Формирование общего языка для оценки систем и процессов

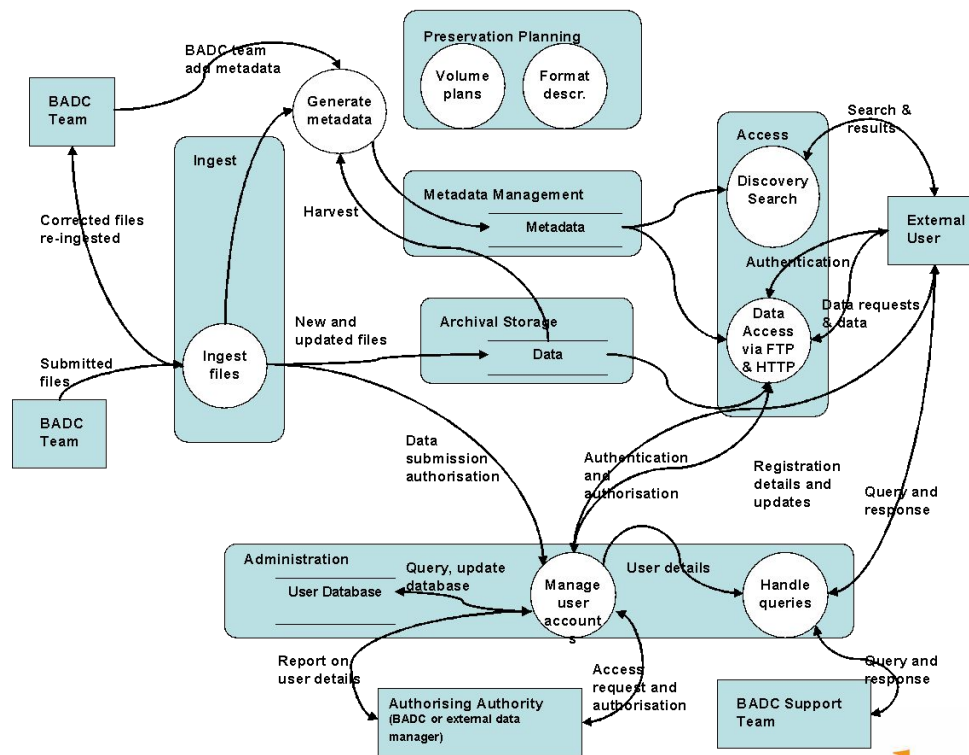
BADC/APS частные исследования (1)

- Британский центр атмосферных данных
 - Центр данных совета исследований по окружающей среде (centre of the Natural Environment Research Council (NERC))
 - Оценка использования сервиса CCLRC's Atlas Petabyte Storage (APS) Service для долгосрочного хранения
 - Привязка модели OAIS к BADC/APS
 - BADC ввод данных и доступ к ним
 - APS архивное хранение
 - Общая ответственность за управление данными и администрирование

BADC/APS частные исследования (2)

- Применение OAIS раскрывает:
 - Обратную связь насколько BADC/APS выполняет требования модели OAIS
 - Выявляет, что AIP нуждается в лучшем определении
 - Слабые стороны при планировании хранения, напр., недостаточный мониторинг технологии или определение группы пользователей
- OAIS помогает идентифицировать ограничения
- Подробнее см.: Corney, *et al.* (2004)
<http://www.allhands.org.uk/2004/proceedings/papers/156.pdf>

BADC/APS частные исследования (3)



Частное исследование UKDA и TNA (1)

- Проект, который финансирует Британский объединенный комитет информационных систем (UK Joint Information Systems Committee (JISC))
- Партнеры:
 - UK Data Archive
 - The National Archives
- Цель связь UKDA and TNA с информационной и функциональной моделями OAIS
- Beedham, *et al.*, *Assessment of UKDA and TNA Compliance with OAIS and METS Standards* (2005)
- <http://www.data-archive.ac.uk/news/publications/oaismets.pdf>

Частное исследование UKDA и TNA(2)

- Заключение:
 - Не существует методологии тестирования соответствия OAIS
 - Рекомендуется создание учебного пособия или руководства
 - Шесть обязательных функций OAIS реализованы почти в любом хорошо организованном архиве
 - Концепт «Определенное сообщество» (Designated Community) OAIS подразумевает идентифицируемую и сравнительно однородную группу пользователей; таковая отсутствует в UKDA лил TNA
 - Связь между AIPs и DIPs нуждаются в прояснении

Частное исследование UKDA и TNA (3)

– Заключение (продолжение):

- Администрирование OAIS может быть трудным для небольших архивов
- Модель не масштабируется
- Категории информации слишком общие чтобы обеспечить соответствие элементов других схем (р. 70)
- НО! Использование терминологии OAIS позволяет поддерживать коммуникацию между UKDA и TNA

Дизайн информирующей системы

Informing system design (1)

- OAIS не «синька» для системы
 - «Подразумевается, что разработчики используют модель как общее руководство" (OAIS 1.4)
- Но модель используется для
 - This can be difficult because the model does not generally distinguish between management and technical processes
 - Need to first identify the areas that can be supported by technical development

Дизайн системы (2)

- Многие примеры:
 - Полные системы:
 - IBM DIAS (used by Koninklijke Bibliotheek)
 - OCLC Digital Archive Service
 - aDORe (Los Alamos National Laboratory)
 - Stanford Digital Repository
 - MathArc (Cornell UL and SUB Göttingen)
 - Инструменты:
 - Программное обеспечение хранилищ: DSpace, FEDORA, ...
 - DCC Representation Information Repository and Registry
 - Harvard University Library XML-based Submission Information Package for e-journal content

Дизайн системы (3)

- Домен специфическое моделирование в качестве ОСНОВЫ
 - Проект InterPARES Preservation Task Force
 - Модель сохранения электронных записей
 - Формальное моделирование специальных процессов и функций связанных с обеспечением сохранности электронных записей
 - Разработка "... спецификации OAIS для различных классов информационных объектов в т.ч. Электронные записи и архивные агрегаты"
 - <http://www.interpares.org/>

Informing system design (4)

- Research projects
 - OAIS is the “guiding principle” of CASPAR
 - CASPAR Conceptual model
 - Representation Information registries and repositories

Сохранность метаданных

Сохранность метаданных

- Метаданные:
 - Данные о данных
 - Структурированная информация об объектах, которая поддерживает разные виды работ: поиск, извлечение, управление и т.д..
 - Часто делятся на описательные, структурные и административные
- Сохранность метаданных
 - Информация в хранилище используется для поддержки процесса цифровой сохранности

Выводы

Выводы

- OAIS хорошо разработан и используется в разнообразных контекстах:
 - Стандартизация терминологии
 - Анализ существующих процессов обеспечения сохранности
 - Дизайн системы (и устройств)
 - Разработка критериев сертификации
 - Дизайн и разработка стандартов метаданных (e.g. PREMIS Data Dictionary)

References

- *Reference Model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1 (2002):
<http://public.ccsds.org/publications/archive/650x0b1.pdf>
- DPC Technology Watch Report on the OAIS model by Brian Lavoie (2004):
http://www.dpconline.org/docs/lavoie_OAIS.pdf
- *Assessment of UKDA and TNA Compliance with OAIS and METS standards* by H. Beedham, et al., (2005):
<http://www.data-archive.ac.uk/news/publications/oaismets.pdf>
- RLG/NARA Task Force on Digital Repository Certification:
http://www.rlg.org/en/page.php?Page_ID=580
- Trusted Repositories Audit & Certification:
<http://www.crl.edu/PDF/trac.pdf>

Ingest exercise

Пример ввода (1)

- Выбрать сценарий:
 - Национальная библиотека создает собрание электронных журналов
 - Университетская библиотека создает собрание публикаций академического персонала
 - Музей или архив осуществляет оцифровку фотографий
 - ...
- Ваш директор спрашивает соответствует ли ваш сервис стандарту OAIS
- Вы детально рассматриваете процесс сохранности и стратегии на предмет соответствия концептам OAIS

Ingest exercise (2)

- For this exercise, we will only consider the Ingest function
 - Ingest is understood as those services and functions that accept SIPs from Producers; prepares AIPs for storage, and ensures that AIPs and their supporting Descriptive Information become established within the OAIS
- Main functions:
 - Pre-Ingest - negotiation and agreement on the nature of SIPs
 - Receive Submission
 - Quality Assurance - for successful transfer
 - Generate AIP - the version stored in Archival Storage
 - Generate Descriptive Information - could be extracted from AIPs
 - Co-ordinate Updates - transfers AIP to Archival Storage and Descriptive Information to Data Management

Упражнение по вводу (3)

- Продумайте требования к определениям пакетов ввода и генерации архивных пакетов
 - Помните, что информационные пакеты больше, чем просто контент – они также включают представление информации
 - Ввод (Ingest) основной интерфейс между OAIIS и создателями контента
 - Создатели контента предъявляют свои требования
 - Уровень «контроля» над создателями может варьировать в зависимости от ситуации
 - Необходимы следующие решения:
 - Какую информацию он собирает (the SIP)
 - Требования к сохранности (the AIP)
 - Как генерировать AIP

Упражнения по вводу данных 4)

- Что следует обдумать:
 - Какие типы объектов вы собираете?
 - Каким образом вы осуществляете сбор объектов?
 - Какие форматы вы используете?
 - Какой уровень контроля у вас над создателями контента?
 - Какой у вас уровень требований к архивному пакету? (существенные свойства)
 - Какое представление информации вам требуется?
 - Какие типы метаданных вам нужны (Preservation Description Information, Descriptive Information) ?
 - Могут ли производители контента снабдить вас метаданными? Если да, как?
 - Как вы упакуете контент в пакеты?

Feedback and discussion

Acknowledgements

- *UKOLN* is funded by the Museums, Libraries and Archives Council, the Joint Information Systems Committee (JISC) of the UK higher and further education funding councils, as well as by project funding from the JISC, the European Union, and other sources. UKOLN also receives support from the University of Bath, where it is based:
<http://www.ukoln.ac.uk/>
- The *Digital Curation Centre* is funded by the Joint Information Systems Committee and the UK Research Councils' e-Science Core Programme:
<http://www.dcc.ac.uk/>

This work is licensed under the Creative Commons Attribution 2.5 Italy License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/2.5/it/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.