

Реинжиниринг данных общего архива наблюдений САО РАН

Желенкова О.П., Витковский В.В., Пляскина Т.А.

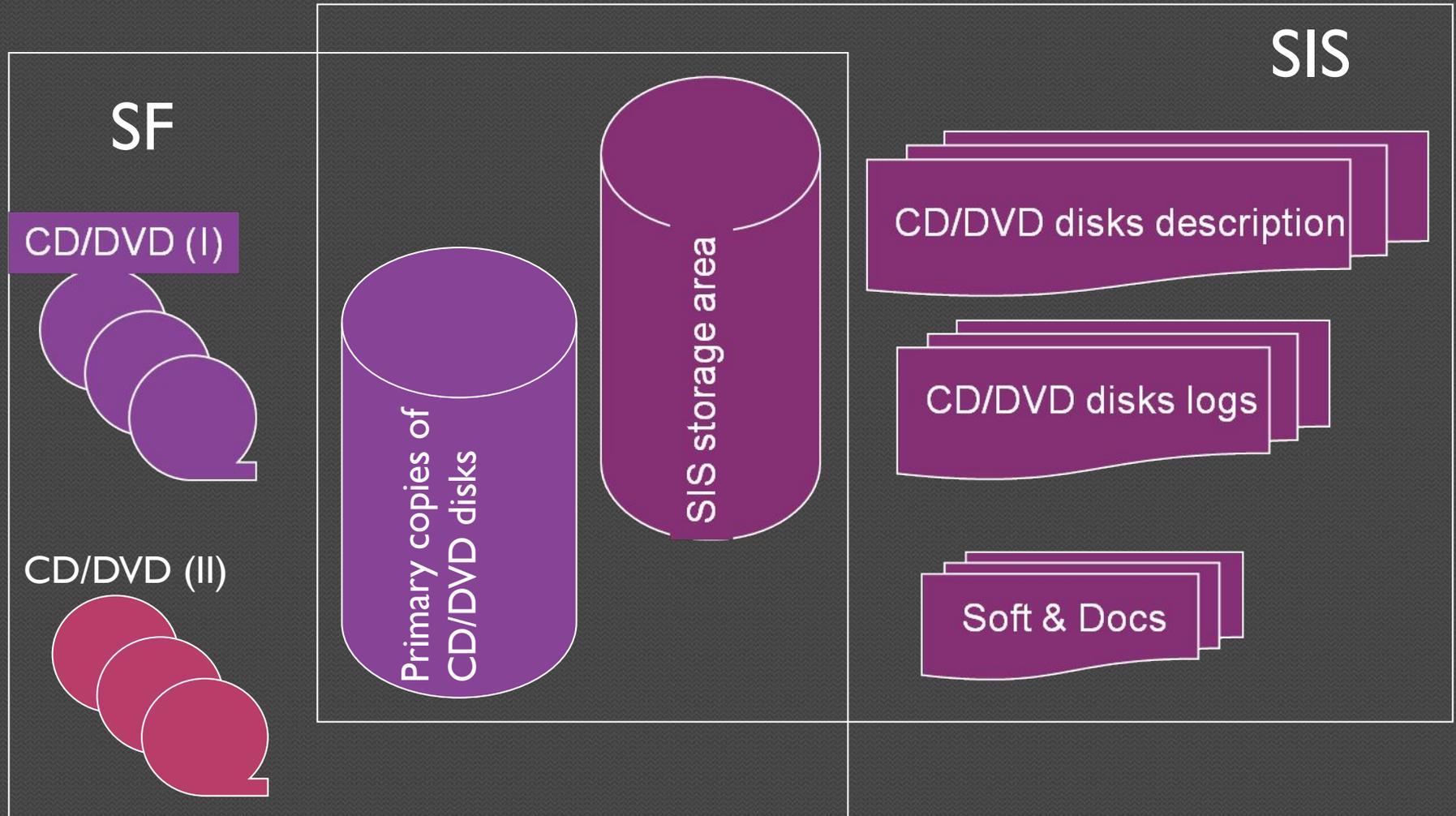
IAU : FITS-формат; классификация астрономических объектов; обозначения объектов; резолюция 5 Комиссии IAU об открытом веб-доступе к цифровым архивам обсерваторий, финансируемых из государственных бюджетов

IVOA : VOTable-формат; Unified Content Descriptors (UCD); протоколы доступа к данным: ConeSearch, SIAP, SSAP, SLAP, TAP ...; Astronomical Data Query Language (ADQL), протоколы для веб-сервисов – PLASTIC, SEA, ...

Требования к организации архивного CD/DVD диска:

- ❖ включает метку диска и каталоги с наблюдениями
- ❖ в каталоге хранятся данные одной ночи
- ❖ название каталога включает дату наблюдений
- ❖ наблюдение хранится в одном файле
- ❖ нет жестких ограничений на формат файла

Архивная система включает хранилище данных (SF) и информационно-поисковую систему (SIS):



Типы архивных файлов (235578):

Наблюдения– 91%

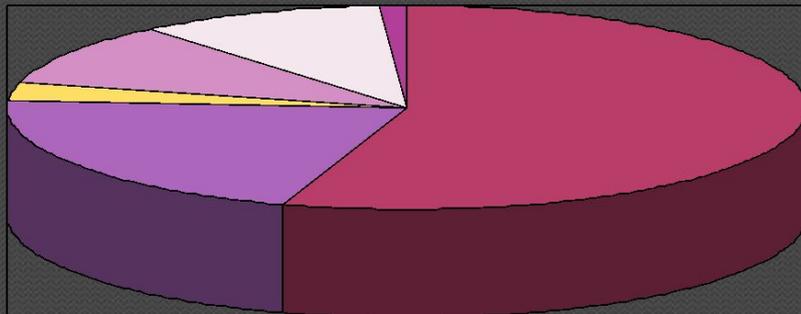
Журналы наблюдений-- 1%

Вспомогательные данные -- 8%

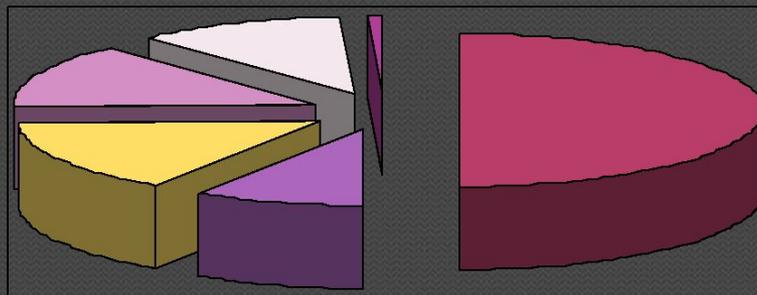
Оптические и радио данные

Архив	CD/DVD диски	Темп прироста данных (МВ)	Объем (GB)	Число записей
Оптика	150+150(копия)	150	309	~190000
Радио	7	4	4	~46000

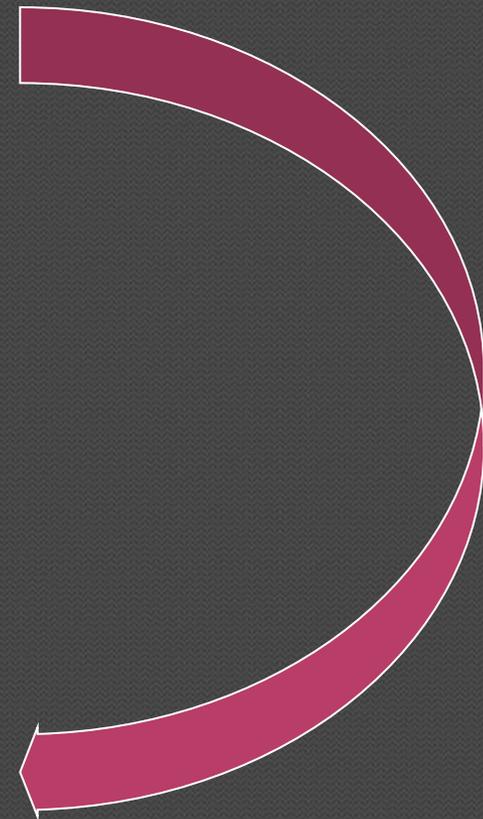
Оптические наблюдательные данные (169890 файлов, 96% - FITS):



- OBJECT - 55%
- BIAS - 21%
- DARK - 3%
- FLAT - 10%
- STANDARD - 10%
- undef - 1%



- DIRECT IMAGE - 55%
- ECHELLE - 10%
| LONG SLIT | 17% |
| MULTI-OBJ | 15% |
| FABRI-PEROT | 12% |
| undef | 1% |



Разделение наблюдательных данных по типу наблюдений:

Прямые снимки	55%
Эшелле-спектры	10%
Длинная щель	17%
Мультиобъектная спектроскопия	15%
Фабри-Перо	12%
Не определяются алгоритмом	<0.1%

Стандартные запросы, кроме выбора данных по дате наблюдения, из-за отсутствия необходимых параметров в заголовках файлов реализуются к части наблюдательных файлов (без учета ошибок в параметрах):

Нет значений координат	17%
Нет имени объекта	4%
Не определено название программы	30%
Не определен заявитель	29%
Не определены наблюдатели	29%

Таблицы поисковой системы

- Таблицы, содержащие информацию для коррекции значений параметров файлов (например: ВТА, 6-m and so on)
- Таблицы и представления с параметрами файлов
- Таблица для связи параметров FITS-файлов и атрибутов поисковой системы



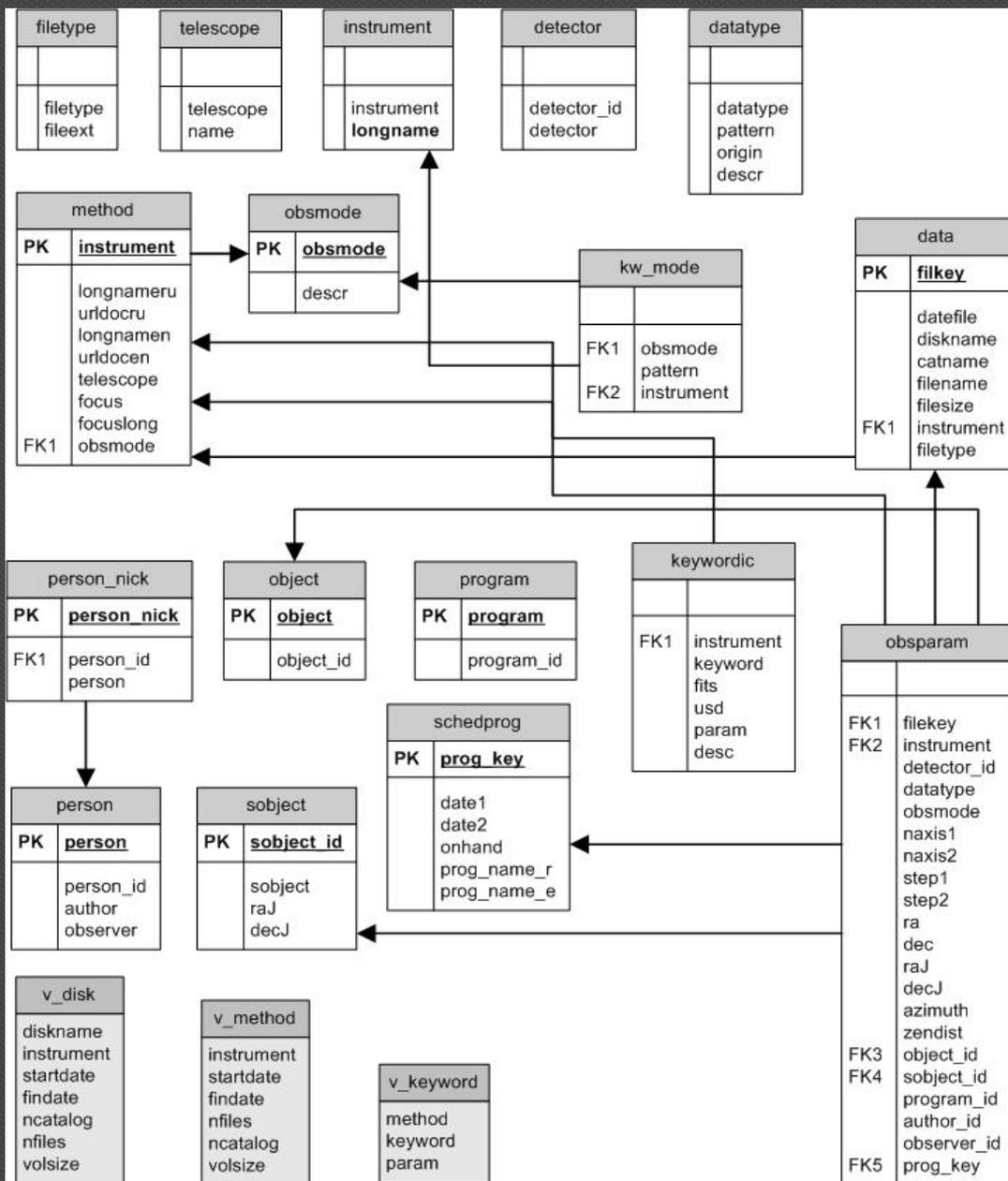
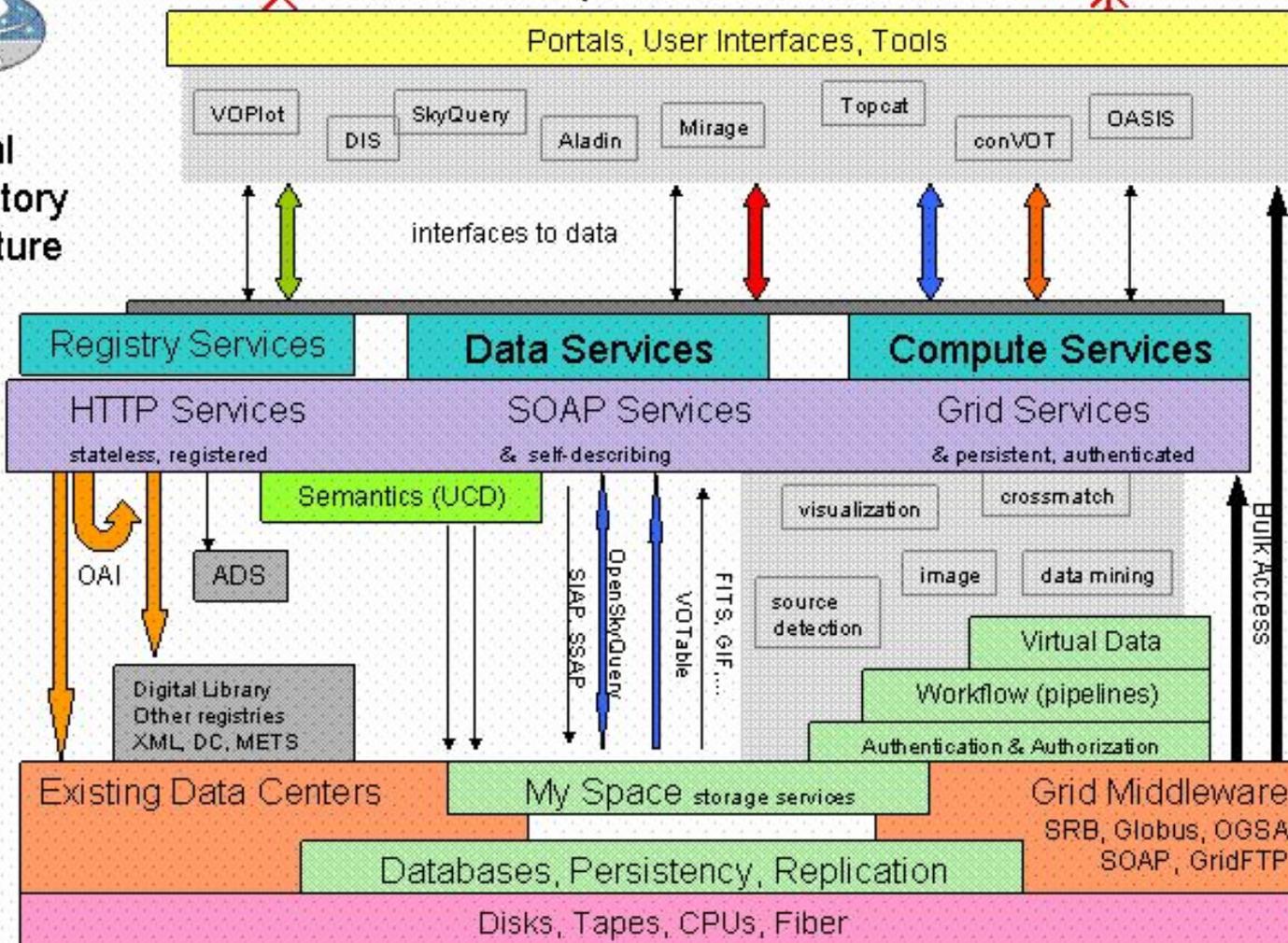


Схема таблиц
ПОИСКОВОЙ СИСТЕМЫ



Virtual Observatory Architecture

Discover Compute Publish Collaborate



Деятельность International Alliance Virtual Observatory (IVOA)

- ❖ регистры
- ❖ модель астрономических данных
- ❖ семантика– Unified Content Descriptors (UCD)
- ❖ доступ к данным
- ❖ язык запросов к астрономическим данным
- ❖ программные сервисы для работы с данными в распределенной среде
- ❖ формат для обмена данными (VOTable)

Сопровождение и хранение (1)

[WG AD, IVOA IG CP]

Определение процедур для сбора информации, необходимой для доступа и управления данными

- оценка – выбор того, что сохранять
- поступление – контролируемый импорт данных
- организация – как структурировать материал
- описание – метаданные для обеспечения подлинности версии, семантического описания и целостности
- хранение - создание архивной формы и помещение в хранилище
- доступ – обнаружение и манипуляции

Сопровождение и хранение (2)

Автоматизированная обработка, как в электронных библиотеках

- Извлечение метаданных
- Регистрирование метаданных и файлов
- Проверка семантики и синтаксиса
- Проверка целостности
- Упаковка данных и метаданных
- Шаблоны потока работ для контроля сопровождения и архивизации
- Независимость от инфраструктуры (постоянные архивы)
 - Стандартные операции для взаимодействия с новой технологией

Спасибо за внимание!

Работа поддержана грантом РФФИ 07-07-00415

