520.27:520.88

Банки наблюдательных данных ODA–R и ODA–B/SS. Фильтры входных оболочек

В. К. Кононов

Специальная астрофизическая обсерватория РАН, п.Нижний Архыз, 369167, Россия

Аннотация. Описываются подсистемы фильтрации входных данных Банков ODA–R и ODA–B/SS. Комплексы программ-фильтров являются частями входных интерфейсных оболочек систем и обеспечивают автоматический контроль и дополнительное преобразование поступающих экспериментальных данных для корректного наполнения стандартизованных архивных баз.

The observational databanks ODA–R and ODA–B/SS. The filters of input shells

V. K. Kononov

Special Astrophysical Observatory of the Russian AS, Nizhnij Arkhyz, 369167, Russia

Abstract. The subsystems of input data filtering for the databanks ODA–R and ODA–B/SS are described. The complexes of the filter programs are the parts of input interface shells of the systems and they provide automatic check and additional conversion of received experimental data for the correct filling of standardized archive bases.

Содержание

1. Введение

Одной из главных задач Банков наблюдательных данных ODA–R (Observational Data Archive – RATAN–600) и ODA–B/SS (Observational Data Archive – BTA/Stellar Spectrographs) является организация получения в режиме “on-line” различных *типов* экспериментальных данных от соответствующих систем сбора телескопов РАТАН–600 и БТА (Кононов и др., 2001; 2002; 2003; Кононов, 2002б). Указанные информационные системы с помощью своего программного обеспечения реализуют непрерывные автоматизированные конвейеры передачи новых данных, которые в итоге помещаются в специализированные разделы систем (процесс архивизации). Разделы Банков данных представляют собой максимальным образом стандартизированные проблемно-ориентированные архивные базы, которые поддерживаются едиными для каждой системы подсистемами управления.

В общем случае степень стандартизации и корректность любых постоянно расширяющихся архивных баз определяется выбором правил межсистемного обмена информацией — между Банками данных и системами сбора. Здесь возможны два противоположных способа взаимодействия. Первый способ возникает в том случае, если каждой системе сбора “предоставляется полная свобода” в выборе форм представления своих выходных данных, конфигураций данных в виде наборов файлов и правил их идентификации. То есть в любой момент времени все эти характеристики могут произвольным образом изменяться в зависимости от специфики наблюдательных программ, методики наблюдений, параметров приемно-измерительных комплексов и даже от человеческого фактора. В результате Банк данных как система попадает в ситуацию полной неопределенности и, по существу, может реализовывать лишь функции элементарной системы буферизации данных на машинных носителях. При этом степень стандартизации архивных баз сводится практически к нулю, а поисковые функции системы будут либо исключены, либо предельно примитивны. Причем все возможные некорректности выходных данных систем сбора будут автоматически переноситься в архивные базы со всеми вытекающими отсюда последствиями.

Второй способ связан с введением строгих соглашений о межсистемном обмене — наборов *локальных* (для каждого типа данных) и *общих* (для любых типов) стандартов. Этот вариант выбран в качестве рабочего для обоих Банков данных. Важно, что наборы стандартов, которые могут учитывать любые специфические особенности информационных выходов отдельных систем сбора на протяжении их суперсетов (Кононов, 1995а), распространяются на 3 области:

- интерпретации данных (конфигурации);
- идентификации данных (именование);
- представления данных (форматы).

Все стандарты формализованы и введены в централизованные описания Банков ODA–R и ODA–B/SS. На этой основе осуществляется полноценный многосторонний контроль входных информационных потоков и как результат — поддержание максимальной степени стандартизации архивных баз и их корректности. В итоге обеспечивается возможность построения различных вариантов подсистем многопараметрического поиска, удовлетворяющих растущие потребности астрономов в этой области. Другими словами, введение стандартов межсистемного взаимодействия позволяет проводить ту или иную предварительную фильтрацию новых данных до их непосредственной архивизации. Роль этого фактора чрезвычайно важна, особенно на фоне интенсивного роста объемов разнородной экспериментальной информации в последние годы.

Данная работа посвящена многоаспектной автоматической фильтрации наблюдательных данных, которая реализована в Банках ODA–R и ODA–B/SS с помощью наборов специализированных программных компонентов, включенных в состав входных интерфейсных оболочек систем.

2. Стандартизация межсистемных связей

В качестве основы стандартизации межсистемного взаимодействия используется унифицированный FLEX-интерфейс (FLexible EXchange), концепция которого была разработана в 1995 г. (Кононов, 1995б). Первоначально вариант этого интерфейса был реализован в Банке ODA–R для данных Rr-типа (радиометры континуума облучателя N 1). В дальнейшем он был распространен и на другие типы наблюдательных данных, включая оптические для Банка ODA–B/SS.

2.1. Поддержка FLEX-интерфейса

Концепция унифицированного интерфейса, которая явилась результатом анализа и обобщения возможных конфигураций межсистемных связей, позволяет реализовать каноническую форму взаимодействия:

$$C_k A P_m \quad (\forall k, m) \quad ,$$

где C_k — системы сбора, A — единая система архивизации (интегрированный Банк данных), P_m — системы обработки. Такая связь может поддерживаться введением стандарта на структуры данных "с двух сторон" в отношении единой архивной системы:

$$D_k = D_0 \longrightarrow D_a \longrightarrow D_0 = D_m \quad (\forall k, m) \quad ,$$

где D_k — выход k-й системы сбора, D_m — вход m-й системы обработки, D_a — множество архивных данных, D_0 — множество стандартизованных данных.

В своей конкретной форме унифицированный FLEX-интерфейс предусматривает набор ограничений, характеризующих общий стандарт взаимодействия систем:

(1) Все системы функционируют в *единой* операционной среде — ОС Linux и размещают множества своих данных в ее файловой системе.

(2) Все системы используют *единый* тип связи между файлами своих множеств данных, соответствующий иерархической модели описателей обобщенной системы сбора.

(3) Все системы используют *единый* внутренний формат для своих файлов, являющийся самодокументируемым на основе специальных дескрипторов.

(4) Все системы используют *единое* правило идентификации своих файлов.

При этом считается, что соответствующие выходы и входы систем сбора и обработки совпадают со стандартным входом/выходом обобщенной системы архивизации:

$$D_k = D_m = D_0 \quad (\forall k, m)$$

и более того, стандарт входа/выхода обобщенной архивной системы распространяется и на сами хранимые данные, то есть $D_a = D_0$.

Для организации взаимодействия с системами сбора вида

$$C_k A \quad (\forall k)$$

Банки данных ODA–R и ODA–B/SS используют одноименные входные интерфейсные оболочки ODA/BS–2L (Oservational Data Archive/Buffering System – 2-Layered) — 2-уровневые сетевые подсистемы буферизации, которые непосредственно реализуют отображения

$$D_k \xrightarrow{A} D_a \quad (\forall k) \quad .$$

Структура и особенности функционирования подсистемы ODA/BS–2L на примере Банка ODA–R подробно описаны в работе Кононова (2002б).

Таким образом, в случае соблюдения стандартов межсистемного взаимодействия процессы передачи и приема новых наблюдательных данных существенно упрощаются и могут полностью контролироваться входными интерфейсными оболочками.

2.2. Перераспределение системных функций

В реальных условиях могут возникать определенные проблемы, когда некоторые системы сбора C_k по тем или иным причинам пока еще не поддерживают принятые соглашения по стандартизации. Это может быть следствием устаревших версий этих систем, модернизация которых в настоящий момент невозможна или нецелесообразна. Для выхода из этого положения и соблюдения идеи стандартизации FLEX-интерфейс предусматривает введение в Банки данных дополнительных компонентов A_k, которые являются частями входных интерфейсных оболочек и предназначены для принудительной стандартизации информационных выходов систем сбора C_k:

$$C_k A_k A \quad (\forall k) \quad .$$

Таким образом могут быть реализованы необходимые отображения по стандартизации структур данных:

$$D_k \neq D_0 \xrightarrow{A_k} D_0 \xrightarrow{A} D_a \quad (\forall k) \quad ,$$

а компоненты A_k будут играть роль дополнительных входных фильтров.

Необходимость введения дополнительных фильтров A_k и перераспределения функций по стандартизации данных между системами сбора C_k и интегрированным Банком данных зависит от текущего состояния и возможностей конкретных систем сбора. При этом важно, что каждый A_k может представлять собой целый набор программ-фильтров, которые выполняются в определенной последовательности.

2.3. Преобразование и контроль данных

В общем случае введение дополнительных подсистем фильтрации данных обусловлено двумя причинами:

1. Необходимостью осуществлять преобразование (в самом широком смысле слова) входных потоков информации.
2. Необходимостью выполнять проверку входных данных на корректность, то есть фиксировать ситуации, когда входные данные по тем или иным критериям отличаются от принятых стандартных соглашений.

Под преобразованием понимается любое изменение входного множества данных D_k, относящегося к определенному временно́му интервалу. Не является преобразованием только лишь пересылка данных из одной области хранения в другую. Примерами преобразования могут служить изменение форматов файлов, переименование файлов, дополнение файлов новой информацией, усечение файлов путем удаления избыточных данных, изменение конфигураций файлов. Простейшим преобразованием считается также дублирование входного множества данных D_k копированием его в другую область хранения (размножение, создание резервных копий).

Контроль входных данных (проверка на корректность) может осуществляться по целому набору правил. Например, проверка на допустимость структур имен файлов, на допустимость конфигураций файлов, относящихся к одному наблюдению, проверка на корректность форматов файлов, на существование необходимых ключевых параметров в дескрипторах файлов и/или на допустимость их значений.

Очевидно, что содержания подсистем фильтрации для разных типов данных могут различаться. Все зависит от требований к уровню контроля входных потоков, степени необходимых преобразований и от сложности структур самих входных данных. При этом отдельные программы-фильтры могут быть в любой момент подключены для обработки потоков входной информации или быть деактивированы.

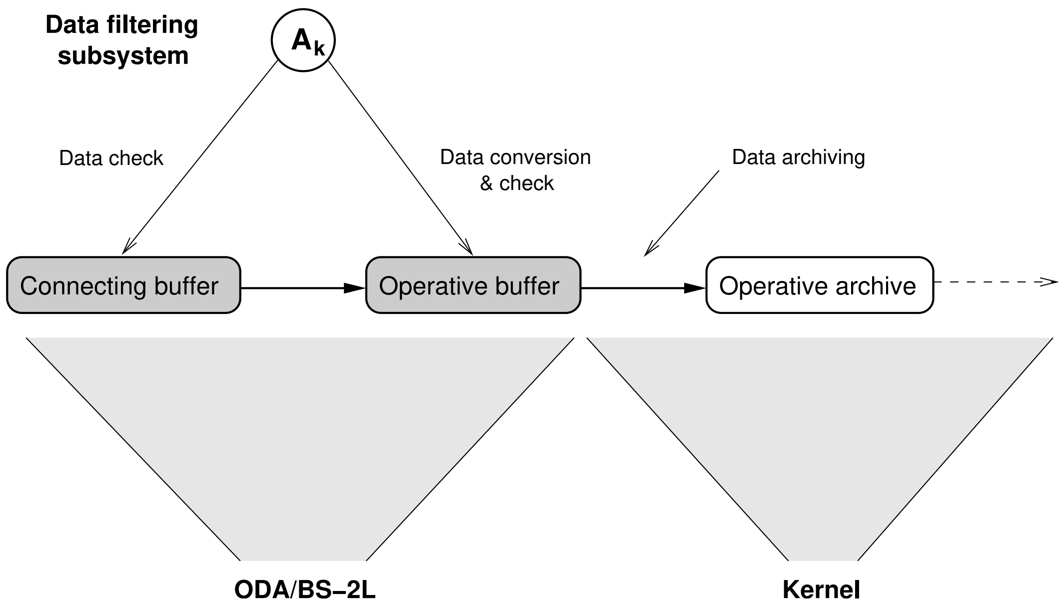

Рис. 1: Области действия подсистем фильтрации A_k для разделов Банков данных ODA–R и ODA–B/SS.

3. Фильтры входных оболочек

Первые варианты законченных в логическом отношении подсистем фильтрации входных информационных потоков были разработаны и введены в эксплуатацию в 2004 г. для 3-х основных типов данных (двух радиоастрономических и одного оптического):

Rr (Continuum)	—	радиометры континуума облучателя N 1 РАТАН–600;
Rs (PAS)	—	панорамный анализатор спектра облучателя N 3 РАТАН–600 (солнечные радиоданные);
Bn (NES)	—	эшелле-спектрограф с коллимированным пучком большого диаметра в фокусе Нэсмит–2 БТА.

В скобках указаны краткие названия приборов.

Выбор этих типов данных был обусловлен их лидерством по объемам экспериментальной информации на радиотелескопе РАТАН–600 и в области звездной спектроскопии на оптическом телескопе БТА.

3.1. Использование подсистемы ODA/BS–2L

Рассмотрим наиболее важные особенности функционирования подсистем фильтрации A_k в рамках входных интерфейсных оболочек. Как было показано в работах Кононова (2002б) и Кононова и Клочковой (2003), главной задачей подсистемы ODA/BS–2L является выполнение подготовительных операций по наполнению *оперативных* буферов разделов Банка новыми данными. В дальнейшем эти данные архивизируются ядром системы посредством перемещения их в соответствующие *оперативные* архивы с регистрацией в справочных базах. Свои выходные данные в виде наборов файлов каждая система сбора C_k регулярно помещает в буфер *связи* раздела, который периодически очищается после передач данных в оперативный буфер (рис. 1).

Любая подсистема фильтрации активируется на 2-х этапах работы оболочки ODA/BS–2L: на первом и на самом последнем. На первом этапе запускается только лишь фильтр контроля данных по именам файлов в буфере связи, чтобы получить информацию о том, какие файлы соответствуют стандарту именования. Именно такие файлы рассматриваются в качестве входного множества данных, которое будет затем проходить все последующие этапы буферизации. Использование фильтра имен на первом этапе имеет принципиальное значение, поскольку может существенно уменьшить интенсивность сетевых обменов между буфером связи и оперативным буфером в случае многомашинной конфигурации (Кононов, 2002б) и большого объема информации.

Основная фильтрация данных выполняется на последнем этапе работы оболочки ODA/BS–2L непосредственно в оперативном буфере. Путем последовательного запуска программ-фильтров анализируются определенные характеристики входных данных и проводятся необходимые преобразования. После выполнения этапов фильтрации в оперативном буфере сохраняются все данные, включая некорректные, но с протоколированием всех возможных замечаний. Такой подход позволяет, с одной стороны, предоставить полную информацию администратору Банка о состоянии новых данных для принятия решений об удалении некорректных данных непосредственно в оперативном буфере или в оперативном архиве уже после их формальной архивизации. С другой стороны, предоставляет возможность максимальным образом архивизировать практически весь полученный уникальный материал несмотря на отдельные отклонения от стандартов.

Логика работы подсистем фильтрации A_k, основанная на анализе и преобразовании данных фактически в одном оперативном буфере, позволяет сколь угодно наращивать состав и возможности комплексов A_k, обеспечивая оптимальные на текущий момент времени процедуры преобразования и контроля информации. Этот факт также подчеркивает некоторую независимость структур выходных данных D_k систем сбора и структур хранимых данных D_a в архивных базах.

Важно подчеркнуть, что отдельные программы-фильтры могут быть взаимосвязанными и требовать определенной последовательности запуска. Это регулируется основными процедурами буферизации типа *tobuf* (Кононов, 2002б) и централизованными описаниями Банков данных.

3.2. Процедуры-фильтры

Процедуры-фильтры представляют собой программы, разработанные на языках *C* и *bash* операционной системы (ОС) Linux. Во втором случае они могут быть оформлены в виде командных файлов или фрагментов командных файлов, выполняющихся в рамках текущего процесса *bash* основных процедур буферизации оболочки ODA/BS–2L. Для каждого типа наблюдательных данных все программы-фильтры вызываются из соответствующей процедуры типа *tobuf*.

3.2.1. Классификация фильтров

В плане оформления и функционирования все программы фильтрации классифицируются на основе следующих критериев (в скобках указан формальный класс):

1. По типу обработки информации:

   (C) Процедуры *контроля* данных выполняют только лишь проверку определенного набора характеристик без какого-либо их изменения.

   (T) Процедуры *преобразования* данных осуществляют изменение информации в соответствии с определенными алгоритмами. Эти преобразования могут касаться любых сторон входных данных как множеств файлов (имен, форматов, структур, конфигураций).

2. По степени универсальности использования:

(U) *Универсальные* процедуры охватывают класс задач, который может выходить за рамки назначения подсистем фильтрации. Такие процедуры могут использоваться другими компонентами (подсистемами) Банка данных для решения своих специфических проблем.

(S) *Функционально-ориентированные* процедуры предназначены только для решения достаточно узкого круга задач фильтрации и в своем максимально развитом варианте могут быть распространены и на несколько типов наблюдательных данных, но для решения одной и той же проблемы.

3. По виду оформления:

(A) *Автономные* процедуры представляют собой отдельно подготовленные в виде модулей программные компоненты, которые запускаются стандартным способом в ОС Linux — с помощью командной строки.

(B) *Встроенные* процедуры являются фрагментами тела основных процедур буферизации типа *tobuf* и выделяются только лишь в связи со своей функциональной ориентацией.

4. По отношению к FLEX-стандартам:

(I) Процедуры поддержки *локальных* FLEX–I_k-стандартов, связанных с конфигурациями входных потоков данных как наборов файлов-описателей модели выхода обобщенной системы сбора (Кононов, 1995а; 1996).

(N) Процедуры поддержки *локальных* FLEX–N_k-стандартов, определяющих правила именования файлов-описателей различных уровней (Кононов, 1995в; 1996; 2002а).

(F) Процедуры поддержки *локальных* FLEX–F_k-стандартов, имеющих отношения к форматам представления файлов на машинных носителях (Верходанов и др., 1994; 1995; Кононов, 1995г).

5. По возможностям динамической активации:

(D) *Отключаемые* процедуры могут быть деактивированы или активированы в любой момент времени, в результате чего степень фильтрации входных данных понижается или повышается.

(V) *Неотключаемые* процедуры не могут быть деактивированы, поскольку их работа является обязательным условием функционирования подсистемы буферизации в целом.

6. По возможностям протоколирования:

(P) *Протоколируемые* процедуры всегда поддерживают специальный журнал, в котором отражаются все возникающие нештатные ситуации (негативные результаты).

(W) *Непротоколируемые* процедуры не предусматривают режим журнализации результатов.

7. По степени взаимозависимости:

(L) *Взаимосвязанные* процедуры могут использоваться только при условии предварительного вызова некоторой другой процедуры (процедур) фильтрации.

(E) *Независимые* процедуры не связаны с результатами работы каких-либо других процедур.

8. По требованию к упорядоченности вызовов:

(R) Последовательность вызовов *упорядоченных* процедур должна быть согласована с фактом активации другой процедуры или процедур.
Процедуры L-класса всегда *упорядочены.*

(X) Последовательность вызовов *неупорядоченных* процедур произвольна.

Принятая и приведенная выше формальная классификация процедур-фильтров представляет более ясную картину текущих возможностей подсистем A_k в целом, что является весьма важным моментом для проведения конкретных программных разработок для повышения эффективности приема и анализа разнородных потоков информации.

3.2.2. Текущие конфигурации

В табл. 1 приведены текущие конфигурации (2004 г.) подсистем фильтрации для 3-х типов наблюдательных данных: Rr, Rs (Банк ODA–R) и Bn (Банк ODA–B/SS). Названия колонок означают:

Data type	—	идентификатор типа данных;
Filter name	—	имя программы-фильтра;
Lang.	—	язык программирования;
Proces. type	—	класс типа обработки;
Univers. charact.	—	класс универсальности;
Design	—	класс оформления;
FLEX	—	класс поддержки FLEX-стандартов;
Activat.	—	класс динамической активации;
Log	—	класс протоколирования результатов;
Link	—	класс взаимозависимости;
Order	—	класс упорядоченности вызовов;
Filter type identifier	—	идентификатор типа фильтра;
Max. dynamic config. ident.	—	идентификатор максимальной динамической конфигурации подсистемы фильтрации.

Особо поясним смысл 2-х последних колонок. *Идентификатор типа фильтра* — это односимвольный идентификатор, который вводится только для фильтров D-класса (динамически активируемых) последнего этапа обработки — обработки в оперативном буфере. Он однозначно определяет конкретный тип программы фильтрации и ее специфику обработки даже при совпадении классов фильтров.

Идентификатор динамической конфигурации представляет собой строку, состоящую из пар символов — знака “+” (признака дополнительного использования фильтра) и идентификатора типа фильтра. Эта строка характеризует, какие в текущий момент времени активированы фильтры D-класса и в какой последовательности. Последняя колонка табл. 1 отражает максимально возможные динамические конфигурации подсистем фильтрации.

Таблица 1: Текущие конфигурации подсистем фильтрации A_k для Банков данных ODA–R и ODA–B/SS

Data type	*Filter name*	*Lang.*	*Classes*								*Filter type identifier*	*Max. dynamic config. ident.*
			Proces. type	*Univers. charact.*	*Design*	*FLEX*	*Activat.*	*Log*	*Link*	*Order*		
Rr	star_	C	C	U	A	N	V	W	E	X	–	+m+s+i
	Rrfin_mirror	bash	T	S	B	–	D	W	E	X	m	
	Rrfin_sizep	bash	C	S	A	F	D	P	E	X	s	
	Rrfin_identp	bash	C	S	A	F	D	P	E	X	i	
Rs	star_	C	C	U	A	N	V	W	E	X	–	+m+n+c+i
	Rsfin_mirror	bash	T	S	B	–	D	W	E	X	m	
	Rsfin_namer	bash	T	S	A	N	D	P	E	R	n	
	Rsfin_conft	bash	C	S	A	I	D	P	L	R	c	
	Rsfin_identp	bash	C	S	A	F	D	P	L	R	i	
Bn	star_	C	C	U	A	N	V	W	E	X	–	+m+s+i
	Bnfin_mirror	bash	T	S	B	–	D	W	E	X	m	
	Bnfin_sizep	bash	C	S	A	F	D	P	E	X	s	
	Bnfin_identp	bash	C	S	A	F	D	P	E	X	i	

Универсальный фильтр *star_* используется во всех подсистемах только на первом этапе — при обработке буферов связи. С его помощью анализируются имена входных файлов на основе гибкой системы *регулярных* выражений (Кононов, 1995д).

Активация фильтра m-типа означает поддержку зеркальной копии оперативного буфера после успешной передачи в него данных из буфера связи. Это повышает надежность временного хранения экспериментальной информации до момента ее полной архивизации ядром Банка.

Фильтры s-типа осуществляют контроль параметров, определяющих размеры входных файлов, представленных в самодокументированном формате. На основе дескрипторов файлов обеспечивается фильтрация всех ситуаций, связанных со сбоями в работе систем сбора, которые проявляются в усечении данных выходных файлов или искусственной избыточности информации.

Фильтры i-типа выполняют контроль наборов ключевых параметров (даты наблюдения, имени объекта и т.п.) в дескрипторах файлов. Наборы таких параметров могут быть разными для различных типов данных, но их корректность определяет правильность формирования справочных баз данных разделов Банков и возможности последующего доступа к архивизированным данным.

Конфигурации подсистем фильтрации для данных типов Rr и Bn подобны. Это связано с подобием соответствующих систем сбора в плане формальной параметризации выходных данных. На облучателе N 3 РАТАН–600 в настоящее время эксплуатируется старая версия системы сбора солнечных радиоданных Rs-типа. С точки зрения “on-line”-поддержки локальных FLEX-стандартов ее возможности пока ограничены. Это проявляется во всех аспектах — в структурах данных, в именовании файлов, в их форматах и наборах файлов. Поэтому до завершения модернизации системы сбора солнечных данных введены два специальных фильтра. Фильтр n-типа осуществляет переименование всех входных файлов, а фильтр c-типа выполняет контроль допустимости конфигураций файлов, относящихся к одному наблюдению. Поскольку контроль данных проводится после преобразования информации в оперативном буфере, подсистема фильтрации данных Rs-типа включает взаимосвязанные и упорядоченные фильтры L и R-класса.

В настоящее время все разработанные подсистемы фильтрации функционируют в своих максимальных рабочих составах, которые отражаются идентификаторами динамических конфигураций в журналах всех основных процедур буферизации типа *tobuf* (Кононов, 2002б; Кононов, Клочкова, 2003). В качестве примера на рис. 2 приведен фрагмент журнала процедуры *Rstobuf_cron* (см. поле *Note*).

3.3. Журнализация нештатных ситуаций

При обнаружении каких-либо некорректностей входных данных процедуры-фильтры P-класса осуществляют запись своих замечаний в специальные журналы ошибок — так называемые E-журналы. Для каждого типа наблюдательных данных поддерживается свой E-журнал, который является одним из протоколов работы соответствующей процедуры типа *tobuf* и может просматриваться ее средствами. Записи в журналах разделяются по фильтрам в той последовательности, как они использовались. При этом фиксируется дата и время обнаружения ошибок, а также приводятся сформированные диагностические сообщения о характере некорректностей. Все журналы регулярно анализируются администратором Банков данных с передачей необходимой информации астрономам-наблюдателям и разработчикам систем сбора.

На рис. 3 и 4 приведены фрагменты реальных журналов ошибок подсистем фильтрации данных соответственно Rs и Bn-типа. Прокомментируем все полученные замечания.

В первом журнале указано, что 21.07.2004 в $16^h00^m38^s$ фильтр *Rsfin_conft*, контролирующий конфигурации входных файлов, обнаружил, что для одного (29-го) наблюдения из полученных 30-ти, нет файла с дополнительными параметрами — файла p-типа. Стандартно

```
# Date      T1      T2     Tsec  Nf    Nkb  Rkbs  RC Knet Kep Klf Kmd  Note
...
20040512  160000  160026   26   12   2420    93   0   6    1   2   1   +m+n+c+i
20040514  140000  140033   33   30   6050   183   0   6    1   2   1   +m+n+c+i
20040514  150000  150025   25    9   1815    73   0   6    1   2   1   +m+n+c+i
20040515  090000  090027   27   15   3025   112   0   6    1   2   1   +m+n+c+i
20040516  100000  100027   27   15   3025   112   0   6    1   2   1   +m+n+c+i
20040517  120000  120423  263   15   3025    12   0   6    1   2   1   +m+n+c+i
20040523  130000  130033   33   33   6605   200   0   6    1   2   1   +m+n+c+i
20040524  110000  110039   39   45   9075   233   0   6    1   2   1   +m+n+c+i
20040609  160000  160031   31   27   5057   163   0   6    1   2   1   +m+n+c+i
20040609  170000  170026   26   12   2420    93   0   6    1   2   1   +m+n+c+i
20040612  110000  110034   34   33   6655   196   0   6    1   2   1   +m+n+c+i
20040615  100000  100037   37   45   9075   245   0   6    1   2   1   +m+n+c+i
20040615  150000  150023   23    3    605    26   0   6    1   2   1   +m+n+c+i
20040616  170000  170027   27   15   3025   112   0   6    1   2   1   +m+n+c+i
20040618  120000  120028   28   15   3025   108   0   6    1   2   1   +m+n+c+i
...
```

Рис. 2: Фрагмент журнала процедуры *Rstobuf_ cron*.

```
...
# Rsfin_conft  20040721  160038
Rs20040721_029_?   h-d
Rsfin_conft: There are incorrect Rs observations (1 of 30) (1)
...
```

Рис. 3: Фрагмент журнала ошибок подсистемы фильтрации данных Rs-типа.

```
...
# Bnfin_identp  20040504  210041
Bn20040504_001o.fits  Obs.type=f
Bn20040504_002o.fits  Obs.type=f
Bn20040504_003o.fits  Obs.type=f
Bn20040504_004o.fits  Obs.type=f
Bnfin_identp: All Bn files are incorrect - 4 (2)

# Bnfin_identp  20040506  190012
Bn20040506_001_.fits  Object=none  Obs.type=?
Bnfin_identp: All Bn files are incorrect - 1 (2)
...
```

Рис. 4: Фрагмент журнала ошибок подсистемы фильтрации данных Bn-типа.

Таблица 2: Статистика обнаружения ошибок подсистемами фильтрации за 2004 г.

Data type	N_d	N_o	N_{oe}	$N_{oe}, \%$
Rr	345	15320	0	0.0
Rs	298	1821	1	0.1
Bn	58	3259	93	2.9

результаты каждого наблюдения с помощью системы сбора солнечных данных должны быть представлены тремя файлами типов h, p и d.

Во втором журнале зафиксировано, что 4.05.2004 в $21^h00^m41^s$ фильтр *Bnfin_identp*, проверяющий корректность ключевых параметров в дескрипторах входных файлов и именах файлов, обнаружил 4 новых файла (наблюдения с 1-го по 4-е), содержащие противоречивые идентификаторы типов наблюдений: в именах файлов — тип “object”, а в заголовках — “flat field”. Затем 6.05.2004 в $19^h00^m12^s$ этот же фильтр обнаружил, что полученный единственный файл не содержит в дескрипторе имя объекта.

3.4. Статистика обнаружения ошибок

Общая статистика обнаружения ошибок подсистемами фильтрации за 2004 г. для трех типов данных приведена в табл. 2. Приняты следующие обозначения:

N_d — число наблюдательных дат;
N_o — число выполненных наблюдений;
N_{oe} — число наблюдений с ошибками;
$N_{oe}, \%$ — относительное число наблюдений с ошибками, %.

Из таблицы следует, что с точки зрения соблюдения локальных FLEX-стандартов система сбора облучателя N 1 РАТАН–600 при круглосуточных наблюдениях в континууме радиодиапазона в течение года (более 15000 наблюдений) не дала ни одного сбоя. Всего лишь один сбой был зафиксирован у системы сбора солнечных радиоданных облучателя N 3 (около 2000 наблюдений в год). В то же время система сбора оптических данных спектрографа NES телескопа БТА передала в выходной поток около 100 некорректных в плане соблюдения стандартов наблюдений. Хотя их относительное число составляет всего лишь 3%, их абсолютное количество при существующих потоках данных Bn-типа эквивалентно одной наблюдательной ночи. При этом все замечания касались параметризации данных: 78% нарушений FLEX–$\mathrm{N_{Bn}}$-стандарта и 22% — FLEX–$\mathrm{F_{Bn}}$-стандарта.

Сравнительный анализ показывает, что фактическое отсутствие некорректностей (в смысле стандартизации) у радиоданных и некорректность оформления результатов каждого 30-го наблюдения, выполненного с помощью спектрографа NES, в первую очередь связаны с различием технологий подготовки и проведения наблюдений на телескопах РАТАН–600 и БТА. На радиотелескопе большинство экспериментов проводится в потоковом режиме по заранее составленным расписаниям, которые дополнительно контролируются Системой подготовки заданий на наблюдения (Жеканис и др., 2002). При этом многие проверенные параметры заимствуются системами сбора из расписаний. На оптическом телескопе существеннее роль человеческого фактора и чаще происходят “on-line”-коррекции наблюдательных программ, учитывающие изменчивость внешних условий проведения наблюдений.

С другой стороны, степень соблюдения стандартных соглашений впрямую зависит от качества построения самих систем сбора как аппаратно-программных комплексов, которые должны адекватно ситуациям (включая сбойные) формализовать свои выходные информационные потоки.

3.5. Реконфигурирование подсистем фильтрации

Реконфигурирование подсистем A_k всегда проводится только администратором Банков данных и может быть связано с 3-мя типами изменений:

- добавлением в подсистему новых программ-фильтров;
- удалением из подсистемы существующих фильтров;
- изменением текущей динамической конфигурации подсистемы.

Первые два момента касаются выполнения определенного объема программных разработок по модернизации основных процедур типа *tobuf*. Поскольку структуры этих процедур подобны, для всех типов данных используется единая технология изменения состава подсистем фильтрации. При этом в большинстве случаев добавление новых компонентов обусловлено необходимостью более жесткого контроля входных данных или ввода дополнительных этапов преобразования, связанных с плановыми изменениями информационных выходов систем сбора. Последнее может также потребовать и исключения ряда уже ненужных программ-фильтров.

Изменение динамических конфигураций подсистем фильтрации (активации тех или иных фильтров D-класса) может рассматриваться как временная модификация, обусловленная текущим уровнем стабильности входных потоков данных, или же как промежуточный этап перед радикальным пересмотром всего состава подсистем A_k (например, в процессе ввода новых версий локальных FLEX-стандартов). Поскольку текущие динамические конфигурации полностью определяются в централизованных описаниях Банков данных, их изменения обеспечиваются лишь простой перенастройкой оболочек ODA/BS–2L по параметрам.

4. Заключение

Разработка и внедрение идеологии фильтрации в режиме “on-line” потоков наблюдательных данных Банками ODA–R и ODA-B/SS означает качественный переход в формировании максимально корректных архивных баз. Единые принципы организации и открытость архитектуры позволяют создавать гибкие по своим возможностям подсистемы фильтрации для различных типов экспериментальной информации. С помощью этих подсистем может быть “сглажена” разнородность существующих систем сбора, которая проявляется как в структурах данных, их форматах и идентификации, так и в степени параметризации наблюдательного материала. Одновременно возникает возможность максимальной стандартизации архивных данных, что имеет принципиальное значение для внедрения гибких унифицированных процедур доступа к большому объему хранимой информации.

Благодарности. Автор выражают благодарность РФФИ за поддержку работы грантами N 02–07–90247 и 02–07–90245.

Список литературы

Верходанов О.В., Кононов В.К., Майорова Е.К., Цыбулев П.Г., 1994, Отчет САО РАН, **233a**

Верходанов О.В., Черненков В.Н., Кононов В.К., Трушкин С.А., Цыбулев П.Г., 1995, Отчет САО РАН, **242**

Жеканис Г.В., Кононов В.К., Мингалиев М.Г., Цыбулев П.Г, 2002, Препринт САО РАН, **167**

Кононов В.К., 1995а, Препринт САО РАН, **110T**

Кононов В.К., 1995б, Препринт САО РАН, **110T**, 13

Кононов В.К., 1995в, Препринт САО РАН, **110T**, 24

Кононов В.К., 1995г, Препринт САО РАН, **111T**

Кононов В.К., 1995д, Препринт САО РАН, **108**, 11

Кононов В.К., 1996, Препринт САО РАН, **114Т**, 15
Кононов В.К., 2002а, Препринт САО РАН, **176**, 14
Кононов В.К., 2002б, Препринт САО РАН, **177**
Кононов В.К., Клочкова В.Г., 2003, Препринт САО РАН, **194**
Кононов В.К., Павлов С.В., Мингалиев М.Г., Верходанов О.В., Нижельская Е.К., Хубиева Н.В., 2001, Препринт САО РАН, **164**
Кононов и др. (Kononov V.K., Pavlov S.V., Mingaliev M.G., Verkhodanov O.V., Nizhelskaya E.K., Khubieva N.V.), 2002, Bull. Spec. Astrophys. Obs., **53**, 131
Кононов и др. (Kononov V.K., Klochkova V.G., Panchuk V.E.), 2003, Bull. Spec. Astrophys. Obs., **56**, 134