

Банк наблюдательных данных ODA–R. Реорганизация основной архивной базы данных радиометров континуума

В. К. Кононов

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

Аннотация. Рассматриваются старая и новая версии форматов архивных томов двух разделов Банка ODA–R — Rr и Rc, содержащих данные радиометров континуума облучателей N1 и 6 радиотелескопа RATAN–600. Описываются процедуры реорганизации основных архивных баз и приводится многолетняя архивная статистика для двух разделов системы.

Observational databank ODA–R. Reorganization of the main archive database of the continuum radiometers

V. K. Kononov

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

Abstract. The old and new versions of the archive volume formats for two divisions Rr and Rc of the Bank ODA–R are considered. These divisions contain the data of the continuum radiometers of feed cabins 1 and 6 of the radio telescope RATAN–600. The procedures of reorganization of the main archive databases are described and the multi-year archive statistics for the two divisions of the system is given.

Содержание

1. Введение	3
2. Основная архивная база Rr-раздела	3
2.1. Форматы архивных томов	3
2.1.1. Формат архивного тома. Версия 1	3
2.1.2. Формат архивного тома. Версия 2	4
2.2. Типы архивов с многотомной организацией	4
2.3. Однородность основной архивной базы	5
2.4. Реорганизация архивных томов	6
2.4.1. Процедуры преобразования	6
2.4.2. Архивная статистика для Rr-данных	7
3. Основная архивная база Rc-раздела	8
3.1. Источник данных для новой архивной базы	9
3.2. Архивная статистика для Rc-данных	10
4. Заключение	11
Список литературы	11

1. Введение

Основная архивная база интегрированного Банка данных ODA-R (Кононов и др., 2001; 2002) представляет собой совокупность архивных файлов, содержащих числовые экспериментальные массивы. Регулярное формирование и поддержка этой базы были начаты в 1989 г. с помощью первой версии системы архивизации ODA Version 1.0 (Кононов, Евангели, 1991). В дальнейшем, в соответствии с концепцией построения интегрированной информационной системы радиотелескопа РАТАН-600 (Кононов, Мингалиев, 1998), основная архивная база была горизонтально фрагментирована по разделам системы, каждый из которых относился к отдельному *minu* наблюдательных радиоданных. При этом большая часть основной базы была вынесена в центральный Rr-раздел Банка ODA-R, включающий данные радиометров континуума облучателей N 1 и 6 (Кононов и др., 1999). На 1.01.2005 г. Rr-раздел системы содержал результаты порядка 188000 наблюдений (около 200000 многочастотных файлов).

В 2005 г., в соответствии с планами развития интегрированного Банка данных ODA-R, были предусмотрены работы по реорганизации основной архивной базы радиометров континуума. Главной целью этих работ являлось создание нового варианта архива, который бы отражал современные требования к хранению большого объема архивной информации и новую методологию доступа пользователей к архивному материалу. Данная статья посвящена различным аспектам этой проблемы, которая в целом была успешно решена.

2. Основная архивная база Rr-раздела

Процесс реорганизации основной архивной базы Rr-раздела коснулся 3-х главных моментов:

- преобразования структуры архивных томов;
- перехода к новой **версии 2** локального FLEX- N_{Rr} -стандарта;
- разделения архивных данных радиометров континуума по облучателям.

Решение всех этих вопросов обеспечило создание нового варианта основной архивной базы с более высокой степенью однородности и более высоким уровнем стандартизации.

2.1. Форматы архивных томов

Две существующие версии формата архивных томов были ранее подробно описаны в работе Кононова (2003). Здесь остановимся только лишь на наиболее важных особенностях этих версий и многолетнем опыте их использования.

2.1.1. Формат архивного тома. Версия 1

Впервые стандартная структура архивного тома в виде формата **версии 1** была введена в 1998 г. для данных Rr-типа при переходе на архивные машинные носители CD (Compact Disk). Это была первая попытка структурировать данные таким образом, чтобы в дальнейшем было удобно осуществлять доступ к информации, используя иерархически связанные каталоги. Формат **версии 1** предусматривал 2-уровневую структуру каталогов — каталоги-месяцы и каталоги-даты. Как показал многолетний опыт работы с этим вариантом формата его главными недостатками оказались следующие:

1. Ориентация на специфику структур данных радиометров континуума, включая принятый в то время локальный FLEX- N_{Rr} -стандарт **версии 0** (Кононов, 2005). Это отразилось на правилах именования каталогов-месяцев и каталогов-дат.

2. Наличие 2-уровневой структуры каталогов с явными ограничениями в идентификации каталогов-месяцев и каталогов-дат приводило к неоправданному усложнению процедур доступа к архивным данным.

3. Отсутствие стандартных правил идентификации архивных томов с помощью файлов-меток ограничивало применение операций автоматического распознавания томов.

Несмотря на перечисленное выше, формат архивных томов **версии 1** был использован для хранения всех архивных данных Rг-раздела, охватывающих 24-летний период 1982–2005 гг.

2.1.2. Формат архивного тома. Версия 2

В 2001 г., в рамках проекта создания интегрированного Банка данных ODA-R, был разработан более удобный и универсальный формат архивных томов **версии 2**. Этот формат предусматривал использование 1-уровневой структуры каталогов — только каталогов-дат. Главными его достоинствами по сравнению с форматом **версии 1** были:

1. Упрощенная схема размещения данных по каталогам и как следствие — более быстрый доступ к архивному материалу.

2. Универсальность правил идентификации каталогов-дат с возможностью более широкого использования для поиска данных регулярных выражений (Кононов, 1995а).

3. Введение стандарта идентификации архивных томов с помощью файлов-меток со структурированными именами.

4. Более развитая самоконтролируемая структура архивных томов, использующая файлы-описатели уровня даты (Кононов, 1995б) со списками контрольных сумм файлов с данными.

Версия 2 формата архивных томов в настоящее время применяется в Банке ODA-R для хранения данных Rs-типа (солнечных радиоданных облучателя N 3) и Ry-типа (спектральных радиоданных облучателя N 2). Этот формат стал также использоваться и для хранения 6-ти типов оптических данных 6-метрового телескопа БТА в Банке ODA-B/SS (Кононов, Панчук, 2001; Кононов, Ключкова, 2003; Кононов и др., 2003). Для соблюдения принципа однородности внутри Rг-раздела архивные тома для данных радиометров континуума до 2005 г. продолжали поддерживаться в формате **версии 1**.

2.2. Типы архивов с многотомной организацией

В настоящее время Rг-раздел системы включает 3 типа архивов, для которых в принципе возможна реализация стандартной структуры архивного тома:

- постоянный архив;
- резидентный архив;
- оперативный архив.

Постоянный архив представляет собой совокупность томов оптических дисков CD-R и предназначен для долговременного эталонного хранения накопленного экспериментального материала. Постоянный архив непрерывно расширяется за счет поступления новых наблюдательных данных и заполнения очередных носителей. На 1.01.2005 г. основная архивная база раздела размещалась на 62-х томах CD типа CDR74 и CDR80 (Кононов, 2003) и включала информацию общим объемом около 40 Gb.

Резидентный архив в общем случае является подмножеством постоянного архива и представляет собой копии томов постоянного архива, созданные на жестком диске HD (Hard Disk) в виде соответствующих каталогов. Главное назначение резидентного архива — это обеспечение быстрого прямого доступа к многолетним архивным данным, минуя операции монтировки томов CD постоянного архива. Текущие ресурсы архивного сервера ПАТАН-600 таковы, что резидентный архив по объему полностью совпадает с постоянным архивом и отражает все изменения последнего.

Таблица 1: Степень однородности основной архивной базы Rr-раздела первой версии

<i>Characteristic</i>	<i>Standard</i>	<i>Degree, %</i>	<i>Change(t)</i>
Archive volume format	Version 1	100	—
Archive file identification	FLEX- N_{Rr} (1)	16	↓
Archive file format	RFLEX(T/R)	100	—
Observation representation	1	96	↑

В отличие от постоянного и резидентного архива *оперативный* архив всегда односторонний. Он размещается на HD (то есть всегда резидентен) и содержит ту часть новых данных основной архивной базы, которая еще не записана на очередной том CD постоянного архива. Обычно объем оперативного архива меньше или немного превышает объем стандартного тома постоянного архива, хотя формального ограничения для этого нет, а фактическое ограничение определяется только лишь свободным пространством на HD.

Для всех 3-х типов архивов поддерживается одинаковый формат архивного тома.

2.3. Однородность основной архивной базы

Понятие *однородности* в широком смысле этого слова может быть связано с различными аспектами построения основной архивной базы. Здесь это понятие используется для оценки состояния архива с точки зрения:

- формата архивных томов;
- правил идентификации архивных файлов;
- формата архивных файлов;
- представления наблюдения набором файлов.

При использовании того или иного критерия для анализа степени однородности из множества доступных должен быть выбран некоторый стандарт (или эталон), который считается более развитым или перспективным в плане универсальности. Относительно этого стандарта можно сделать количественные оценки для тех объектов архивной базы, которые соответствуют выбранному критерию. Например, для первого критерия из приведенного выше списка таким объектом является архивный том, для второго и третьего — архивный файл, для четвертого — наблюдение. Полученные таким образом значения и будут характеризовать степень однородности основной архивной базы в различных аспектах.

В табл. 1 приведены оценки степени однородности основной архивной базы Rr-раздела на 1.01.2005 г. до ее реорганизации (первая версия раздела). Поясним содержание таблицы.

С точки зрения формата архивных томов основная архивная база всегда была на 100% однородной. Это было обеспечено благодаря целенаправленной поддержке формата **версии 1**, несмотря на введение для других типов данных более совершенной **версии 2**.

Однородность в плане стандартной идентификации архивных файлов составила всего лишь 16%, поскольку эталоном для оценки считается **версия 1** локального FLEX- N_{Rr} -стандарта, которая была использована для 31133-х файлов восстановленных старых наблюдений (Кононов, 2005). При этом система сбора облучателя N 1 продолжала создавать выходные файлы в устаревшем локальном стандарте идентификации **версии 0**. По этой причине степень однородности архивной базы со временем постоянно снижалась (колонка *Change(t)*).

Однородность по части формата архивных файлов была стабильна и всегда держалась на уровне 100%, так как все свои выходные файлы система сбора формировала в стандартном регистрационном RFLEX-формате (вариант RFLEX-R). Восстановленные старые данные бы-

ли также приведены к RFLEX-формату (транспортный вариант RFLEX-T) (Кононов и др., 1998; Кононов, Павлов, 1999). Отличия вариантов форматов можно игнорировать, поскольку RFLEX-формат является самодокументированным.

Степень однородности архивной базы в плане представления наблюдения набором файлов составила 96%. Это связано тем, что стандартом считается “одно наблюдение – один многочастотный файл”, в то время как в период подключения новых радиометрических каналов в 1996–1997 гг. системой сбора облучателя N 1 7518 наблюдений были представлены 2–3-мя файлами. Так как число таких “нестандартных” наблюдений фиксировано, то с ростом объема архивной информации степень однородности архивной базы по этому критерию постоянно повышается.

Таким образом, главное, на что следовало обратить внимание в процессе реорганизации основной архивной базы радиометров континуума, было падение степени ее однородности из-за продолжения использования локального FLEX-N_{Rr}-стандарта **версии 0**.

2.4. Реорганизация архивных томов

Главными целями преобразования архива данных Rr-типа и создания его новой версии являлись:

- переход к локальному FLEX-N_{Rr}-стандарту **версии 2**;
- переход к формату архивных томов **версии 2**;
- удаление из архива данных, не относящихся к облучателю N 1.

2.4.1. Процедуры преобразования

В качестве источника данных для проведения реорганизации основной архивной базы был использован резидентный архив Rr-раздела с его многотомной структурой. Преобразование осуществлялось по принципу “том – в том”. Для выполнения максимально автоматизированной конверсии архивных томов был разработан специализированный комплекс программ и командных файлов, основу которого составили две процедуры – *Rrv12_fn02* и *Rrv12_fn12*.

Процедура *Rrv12_fn02* обеспечивает преобразование вида

$$\left\{ \begin{array}{l} \text{Volume format (1)} \\ \text{FLEX-N}_{Rr} (0) \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Volume format (2)} \\ \text{FLEX-N}_{Rr} (2) \end{array} \right\} ,$$

а процедура *Rrv12_fn12* реализует

$$\left\{ \begin{array}{l} \text{Volume format (1)} \\ \text{FLEX-N}_{Rr} (1) \end{array} \right\} \longrightarrow \left\{ \begin{array}{l} \text{Volume format (2)} \\ \text{FLEX-N}_{Rr} (2) \end{array} \right\} .$$

Таким образом, главным отличием процедур является то, что первая из них ориентируется на входные файлы, соответствующие локальному FLEX-N_{Rr}-стандарту **версии 0**, вторая – **версии 1**. При этом для конверсии исходных стандартов в стандарт **версии 2** процедуры используют соответственно базовые программы *RFL02rename_* и *RFL12rename_*, описанные в работе Кононова (2005).

Особенностями перехода от одного локального стандарта к другому являлись:

1. Преобразование исходных имен объектов на основе таблицы перекодировки из централизованных описаний Банка ODA-R: удаление из исходных имен символов *, ?, \$, \, /, ", ', ' и управляющих символов и замена символа “пробел” на _.

2. Замена исходных пустых имен объектов на стандартное имя из централизованных описаний – односимвольное имя _.

```

Default object name - "_"
Object name conversion table:
  soh del  sp _ " del  ' del  ' del  / del  \ del  $ del  * del  ? del

20040116:  files renamed - 30
20040117:  files renamed -  9
20040118:  files renamed - 29
20040119:  files renamed - 20
20040120:  files renamed - 30
20040121:  files renamed - 30
20040122:  files renamed - 30
20040123:  files renamed - 30
20040124:  files renamed - 27
20040125:  files renamed - 28
...

Md5 sum calculation... done

Dates processed      -      19
Files renamed        -     487
New volume capacity -     697 Mb

```

Рис. 1: Фрагмент протокола процедуры *Rrv12_fn02* при обработке отдельного тома.

3. Установка в выходных именах модификатора M из множества символов $\{a, b, \dots, z\}$, если исходные наблюдения представляются более чем одним многочастотным файлом.

Для контроля операций конверсии архивных томов все этапы работ процедур *Rrv12_fn02* и *Rrv12_fn12* протоколировались. На рис. 1 приведен пример фрагмента протокола процедуры *Rrv12_fn02* при обработке отдельного тома.

В результате выполненных преобразований и последующего удаления данных облучателя N 6 была создана новая версия *резидентного* архива Rr-раздела, включающая на 1.01.2006 г. 76 образов архивных томов общим объемом 48.4 Gb. На ее основе был сформирован *постоянный* архив Rr-данных на томах CD типа CDR80.

2.4.2. Архивная статистика для Rr-данных

Здесь, а также в разд. 3.2, используются следующие обозначения:

Таблица 2: Общая статистика по реорганизованной основной архивной базе Rr-раздела

<i>Year</i>	N_d	N_o	N_f	N_{Mb}	<i>Year</i>	N_d	N_o	N_f	N_{Mb}
1982	8	8	8	0	1994	273	2723	2723	217
1983	4	4	4	0	1995	279	8950	8950	651
1984	47	222	222	7	1996	327	17054	19574	1150
1985	116	474	474	14	1997	311	14020	23186	1486
1986	106	870	870	28	1998	316	10195	10195	1889
1987	207	2949	2949	53	1999	343	14825	14825	3641
1988	238	3758	3758	59	2000	342	14232	14232	5102
1989	266	4486	4486	97	2001	357	20370	20370	5056
1990	334	5173	5173	80	2002	334	22372	22372	5485
1991	324	4443	4443	112	2003	348	22004	22004	8189
1992	52	1779	1779	12	2004	345	15320	15320	6400
1993	115	1477	1477	123	2005	353	20912	20912	9730
<i>Total</i>						5745	208620	220306	49581

N_d	—	число наблюдательных дат;
N_o	—	число наблюдений;
N_f	—	число архивных файлов;
N_{Mb}	—	объем данных, Mb;
N_{Gb}	—	объем данных, Gb;
M_{od}	—	среднее число наблюдений на дату;
Max_{od}	—	максимальное число наблюдений на дату;
M_{Mbd}	—	средний объем суточных данных, Mb;
M_{kbd}	—	средний объем суточных данных, kb;
Max_{Mbd}	—	максимальный объем суточных данных, Mb;
Max_{kbd}	—	максимальный объем суточных данных, kb;
M_{co}	—	среднее число каналов на наблюдение;
Max_{co}	—	максимальное число каналов на наблюдение;
M_{kbf}	—	средний размер файла, kb;
Max_{kbf}	—	максимальный размер файла, kb.

В табл. 2 приведена общая статистика по реорганизованной основной архивной базе Rr-раздела на 1.01.2006 г., в табл. 3 — статистика уровня даты, наблюдения и файла.

Значения параметров N_{Mb} , M_{Mbd} и Max_{Mbd} указаны с учетом объемов каталогов-дат, предусмотренных форматом архивных томов **версии 2**.

Рис. 2 иллюстрирует распределение по годам объема данных Rr-раздела за весь архивный период.

3. Основная архивная база Rс-раздела

Решение о формировании нового Rс-раздела было принято в 2005 г. на первом этапе выполнения проекта развития интегрированного Банка ODA-R. Оно было связано с желанием разделить экспериментальный материал, полученный с помощью отличающихся приемно-измерительных комплексов радиометров континуума облучателей N 1 и 6. По этой причине был введен новый формальный *тип* данных Rс, относящийся только к облучателю N 6 (Кононов, 2005).

Таблица 3: Статистика уровня даты, наблюдения и файла для реорганизованной основной архивной базы Rr-раздела

<i>Year</i>	M_{od}	Max_{od}	M_{Mbd}	Max_{Mbd}	M_{co}	Max_{co}	M_{kbf}	Max_{kbf}
1982	1	1	0	0	4	5	7	15
1983	1	1	0	0	3	3	16	24
1984	5	22	0	2	4	6	28	94
1985	4	18	0	0	5	6	26	95
1986	8	27	0	1	3	6	29	115
1987	14	68	0	1	5	7	16	115
1988	16	73	0	1	5	9	15	70
1989	17	62	0	2	5	7	20	109
1990	15	53	0	1	4	7	14	28
1991	14	52	0	5	3	7	24	348
1992	34	53	0	1	3	6	5	46
1993	13	25	1	3	4	6	79	143
1994	10	56	1	4	5	6	76	163
1995	32	67	2	15	6	10	72	635
1996	52	120	4	12	9	14	56	898
1997	45	87	5	22	8	15	62	6329
1998	32	87	6	16	16	19	184	1565
1999	43	96	11	45	17	27	246	1860
2000	42	97	15	43	25	27	360	3806
2001	57	169	14	41	21	27	248	4989
2002	67	181	16	46	19	24	245	2474
2003	63	192	24	49	16	17	375	4056
2004	44	113	19	64	17	29	422	2612
2005	59	189	28	61	17	19	470	2612
<i>Total</i>		192		64		29		6329

3.1. Источник данных для новой архивной базы

Экспериментальный материал от радиометров континуума облучателя N 6 был получен только лишь в 1988 г. в результате проведения серии коротких циклов наблюдений и исторически был включен в состав Rr-архива. Несмотря на сравнительно небольшой объем, эти данные имеют важное самостоятельное значение как в методическом, так и в астрофизическом плане, поскольку были зарегистрированы при наблюдениях с использованием всего Кругового отражателя антенны РАТАН-600 (Верходанов, 1993; Парийский и др., 1994). В связи с этим источником информации для формирования основной архивной базы нового Rc-раздела послужил вариант *резидентного* Rr-архива на этапе его реорганизации.

Данные Rc-типа были автоматически идентифицированы процедурой *Rrv12_fn12* в процессе конверсии архивных томов. Далее на их основе был создан однотомный резидентный Rc-архив и соответствующий том постоянного архива на CD типа CDR24. Архивные тома имеют формат **версии 2**, а имена файлов соответствуют **версии 2** локального FLEX-N_{Rc}-стандарта (Кононов, 2005). Сформированная таким образом архивная база включает Rc-данные с условным статусом “старые”. Однако она имеет потенциальную возможность расширяться в случае введения в эксплуатацию на облучателе N 6 нового радиометрического комплекса.

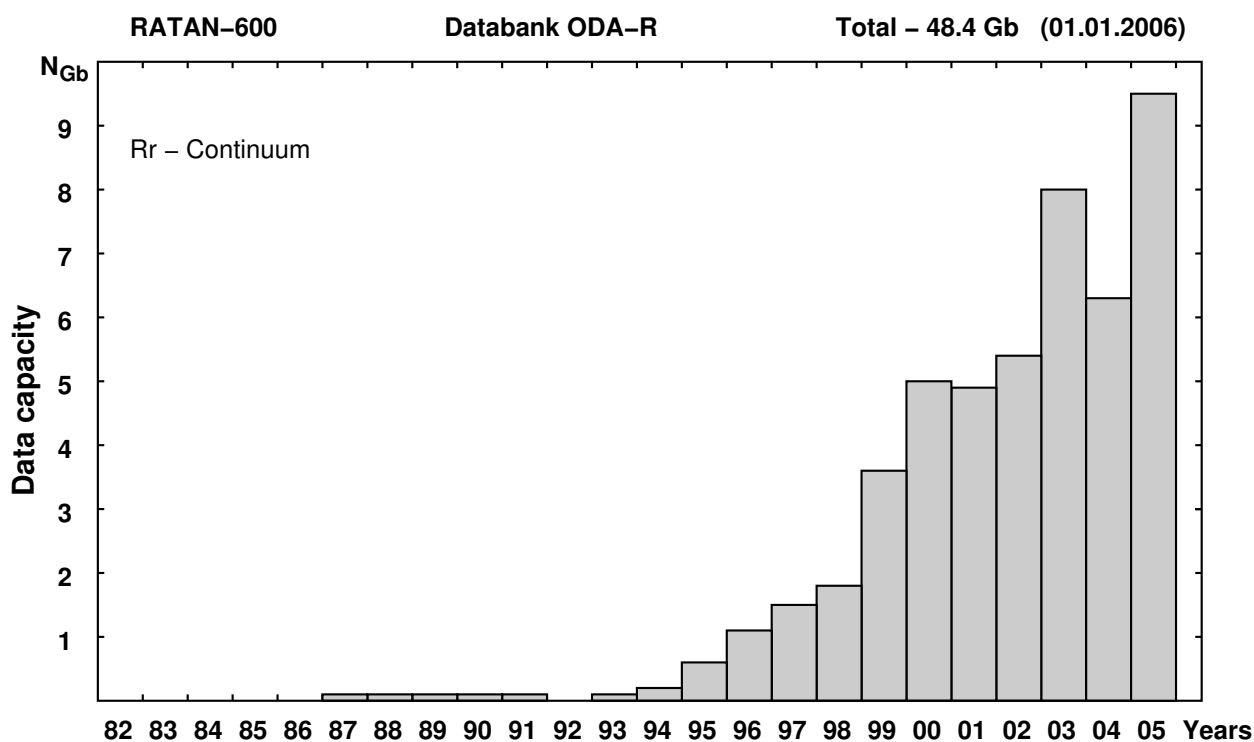


Рис. 2: Распределение по годам объема данных Rr-раздела Банка ODA-R.

Таблица 4: Общая статистика по созданной основной архивной базе Rс-раздела

Year	N_d	N_o	N_f	N_{Mb}
1988	68	438	438	6

3.2. Архивная статистика для Rс-данных

В табл. 4 приведена общая статистика по созданной основной архивной базе Rс-раздела на 1.01.2006 г., в табл. 5 — статистика уровня даты, наблюдения и файла.

Значения параметров N_{Mb} , M_{kbd} и Max_{kbd} указаны с учетом объемов каталогов-дат, предусмотренных форматом архивных томов **версии 2**. Параметры M_{co} и Max_{co} имеют значения 1, так как все наблюдения проводились в одноканальном режиме на длине волны 8 см.

Таблица 5: Статистика уровня даты, наблюдения и файла для созданной основной архивной базы Rс-раздела

Year	M_{od}	Max_{od}	M_{kbd}	Max_{kbd}	M_{co}	Max_{co}	M_{kbf}	Max_{kbf}
1988	6	14	97	196	1	1	11	13

4. Заключение

В результате реорганизации основной архивной базы Rг-раздела был сформирован новый вариант *резидентного* архива, содержащего на 1.01.2006 г. образы 76-ти томов общим объемом 48.4 Gb и включающего 208620 многочастотных наблюдений (220306 файлов). На завершающем этапе на основе резидентного архива был создан *постоянный* архив на томах CD типа CDR80 (эталон и копия). Одновременно было проведено разделение наблюдательных данных облучателей N 1 и 6 и образован новый Rс-раздел Банка данных ODA-R со структурой, подобной Rг-разделу.

Таким образом, к началу 2006 г. многолетняя архивная база данных радиометров континуума РАТАН-600 была приведена к максимально однородному виду. Это обеспечило возможность начать разработку качественно новой унифицированной подсистемы многопараметрического поиска архивного материала.

Благодарности. Автор выражает благодарность РФФИ за поддержку работы грантом N 05-07-90084. Автор также признателен Л.В. Минаковой за помощь в подготовке текста данной статьи.

Список литературы

- Верходанов О.В., 1993, Методы исследования радиоисточников в режиме околосенитного синтеза на РАТАН-600. — Кандидатская диссертация, САО РАН, 192с.
- Кононов В.К., 1995а, Препринт САО РАН, **108**, 11
- Кононов В.К., 1995б, Препринт САО РАН, **110Т**
- Кононов В.К., 2003, Препринт САО РАН, **191**
- Кононов В.К., 2005, Препринт САО РАН, **207**
- Кононов В.К., Евангели А.Н., 1991, Сообщ. САО, **67**, 87
- Кононов В.К., Мингалиев М.Г., 1998, Препринт САО РАН, **129Т**
- Кононов В.К., Павлов С.В., Мингалиев М.Г., Верходанов О.В., 1998, Препринт САО РАН, **128Т**
- Кононов В.К., Павлов С.В., 1999, Препринт САО РАН, **130Т**
- Кононов В.К., Павлов С.В., Мингалиев М.Г., Верходанов О.В., 1999, Препринт САО РАН, **131Т**
- Кононов В.К., Панчук В.Е., 2001, Препринт САО РАН, **163**
- 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 В.К., Клочкова В.Г., 2003, Препринт САО РАН, **194**
- Кононов и др. (Kononov V.K., Klochkova V.G., Panchuk V.E.), 2003, Bull. Spec. Astrophys. Obs., **56**, 134
- Парийский и др. (Parijskij Yu.N., Verkhodanov O.V., Panchuk G.A., Bajkova A.T., Golnev A.Ya., Esepkina N.A., Zverev Yu.N., Majorova E.K., Mingaliev M.G., Opeikina L.V., Stotskij A.A., Temirova A.V., Fridman P.A.), 1994, Bull. Spec. Astrophys. Obs., **38**, 58