

# Задачи в области звездной астрономии и поддержка их данными

О.Ю.Малков, Н.А.Скворцов,  
Е.А.Аввакумова, О.Б.Длужневская,  
А.А.Исаева, П.В.Кайгородов,  
Л.А.Калиниченко, С.В.Карпов, Е.Ю.Кильпио,  
Д.А.Ковалева, А.В.Миронов, С.Г.Сичевский,  
Д.А.Чулков

Институт астрономии РАН  
Физический факультет МГУ им. Ломоносова  
Кафедра астрономии и геодезии УрФУ  
Астрокосмический центр ФИАН  
Специальная астрофизическая обсерватория РАН  
Институт проблем информатики РАН  
Государственный астрономический институт им. Штернберга МГУ им. Ломоносова

RCDL-2014, Ярославль, Октябрь 2014

# План

- Новая парадигма развития и науки с интенсивным использованием данных,
- Тенденции, приведшие к необходимости создания средств Виртуальной обсерватории и международного альянса, сложившегося для её развития и эксплуатации в научных исследованиях.
- Задачи, решённые в области звёздной астрономии с использованием средств Виртуальной обсерватории, а также подход к решению общих задач, возникающих у большинства исследователей.
- Перспективы развития и актуальные проблемы Виртуальной обсерватории.

# Звездная астрономия

- Звездная астрономия (называемая также галактической астрономией) – раздел астрономии, исследующий нашу Галактику и, в частности, строение, динамику, образование и эволюцию входящих в нее звездных систем. Традиционно для решения большинства задач звездной астрономии привлекаются большие массивы наблюдательных и теоретических данных.

# Четвертая парадигма

- Четвёртая парадигма научных исследований сегодня прорисовывается всё отчетливее. В разные времена в основе развития науки лежали, **во-первых**, натуральный эксперимент и установление фактов, **во-вторых**, развитие теорий и их проверка, **в-третьих**, применение вычислительных методов и моделирования. **Четвёртая** парадигма предполагает исследование открытых массивов научных данных больших объемов.

# От поиска данных – к анализу

- Происходит переход от поиска данных для решения задач к всестороннему анализу массивов данных имеющимся арсеналом научных методов и инструментов для выявления полезной и новой информации и знаний. Исследование начинается со сбора данных от инструментов наблюдения или моделирования, возможно не связанных с исследователем.
- Добытые данные в процессе курирования подвергаются первичной обработке, структурированию и классификации, семантическому анализу в соответствии со знаниями о предметных областях исследования. Они снабжаются информацией об их происхождении, обрабатываются для получения вторичной информации методами, имеющимися в предметной области, и представляются в форме, эффективной для долговременного хранения, открытого доступа в исследовательских средах и дальнейшего анализа.

# Вал наблюдательных данных

- В связи со значительными достижениями в наблюдательных, регистрирующих и вычислительных технологиях астрономия в конце прошлого века столкнулась с лавинообразным увеличением количества получаемых данных. Эти наборы данных покрывают небо в различных диапазонах длин волн – от гамма- и рентгеновского диапазона, через ультрафиолетовый, оптический и инфракрасный до радиодиапазона. Для решения многих астрофизических задач накопленного сейчас наблюдательного материала уже достаточно. С учётом появления недорогих технологий хранения данных и наличия высокоскоростных сетей концепция мульти-терабайтных, бесшовных интероперабельных баз данных уже не считается надуманной.
- Всё большее количество астрономических каталогов становятся взаимосвязанными, поисковые машины всё более и более усложняются, а результаты анализа таких данных оказываются столь же богатыми, как и для данных, полученных с реальных телескопов. Предназначенные для обзоров телескопы наземного и космического базирования могут поставлять изображение всего неба за несколько дней и производить, таким образом, данные в объёмах, измеряемых в петабайтах.

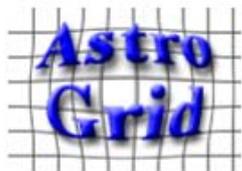
# Виртуальная обсерватория

- За прошедшие годы концепция виртуальной обсерватории, призванной удовлетворять существующим требованиям к управлению данными, их анализу и распространению, завоевала широкую популярность. Это система, в которой распределённые по всему миру гигантские астрономические архивы и базы данных интегрированы в единую среду вместе с инструментами анализа и вычислительным сервисом.
- Необходимо отметить, что виртуальная обсерватория является эффективным инструментом «демократизации» астрономии, поскольку даже небольшие научные, а также образовательные астрономические учреждения получают, по существу, те же возможности для проведения научных исследований на современном наблюдательном материале, что и ведущие астрономические организации мира.

# Электронная наука

- Виртуальная обсерватория представляет собой механизм реализации концепции электронной науки в астрономии. Говоря упрощённо, увеличение научного выхода данных означает получение большего количества научных результатов (например, опубликованных статей или докладов, представленных на конференциях) с каждого гигабайта данных, приходящих с данного конкретного инструмента.

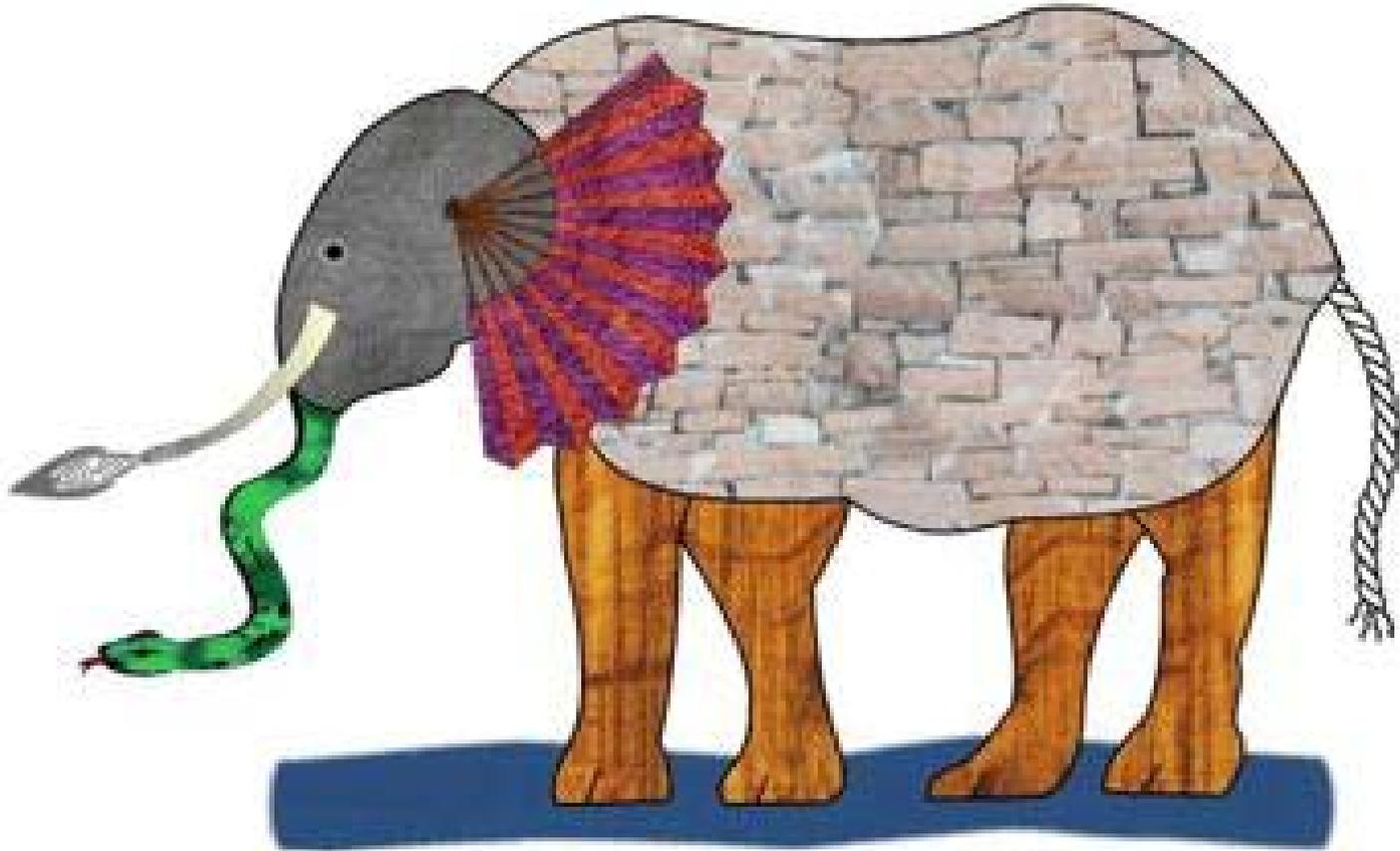
# International Virtual Observatory Alliance ([www.ivoa.net](http://www.ivoa.net))



# Мультихроматический мультивременной подход



# Мультихроматический мультивременной подход



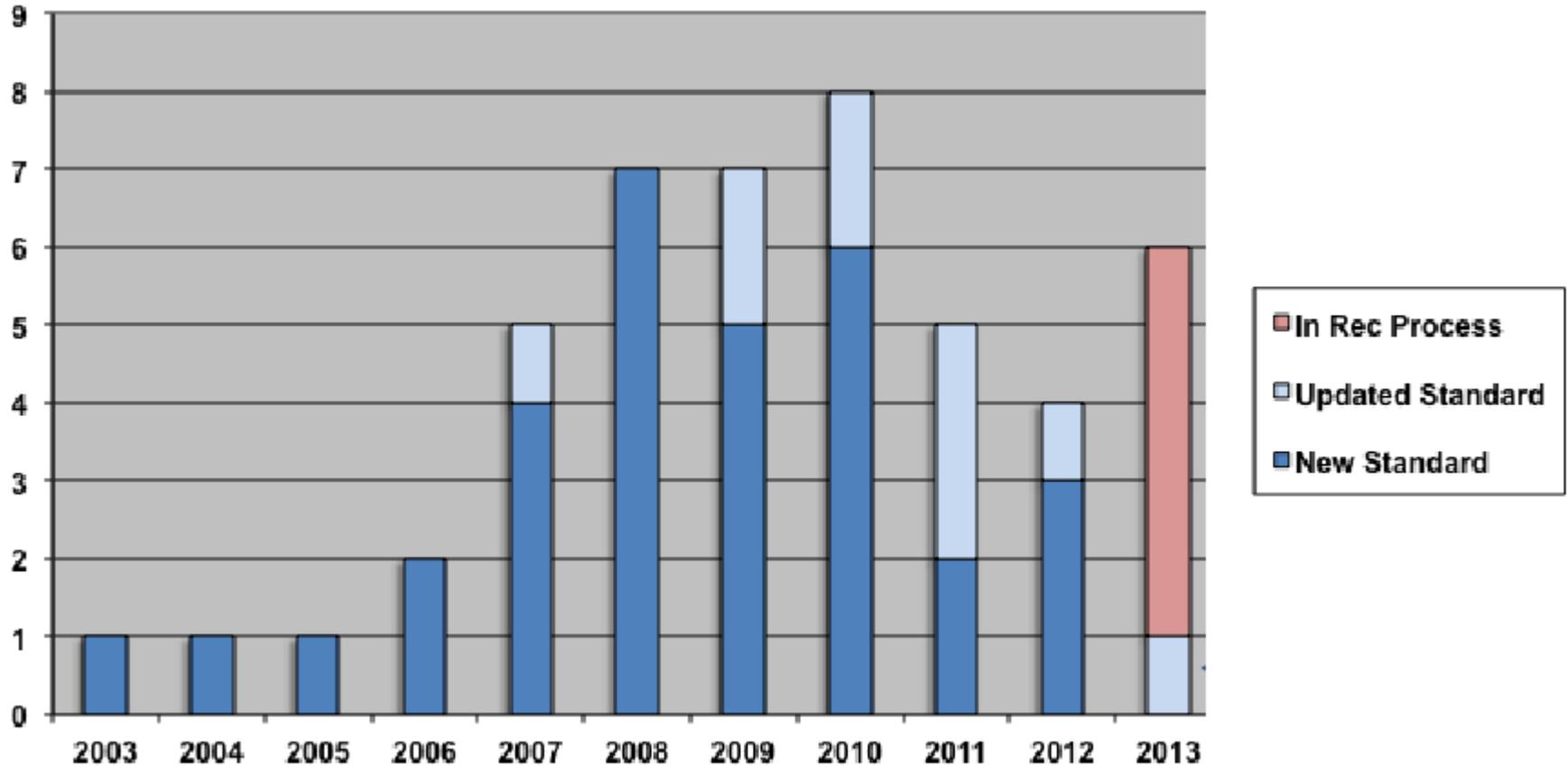
... должен сопровождаться разработкой стандартов

# Стандарты и рекомендации IVOA

- Представление и формализация данных
  - Протоколы доступа к данным и обмена сообщениями для программных приложений
  - Веб-сервисы и грид-сервисы
  - Описание и публикация ресурсов
  - Язык запросов
  - Поддержка сохранности данных
- ..... > 30 утвержденных стандартов

# Стандарты IVOA

IVOA Standards issued per year



# Пример формата, одобренного IAU: Flexible Image Transport Format

- FITS – самый популярный формат для обмена данными в астрономии. Одобрен IAU. FITS используется для передачи и хранения изображений (и табличных данных).
- Ватикан принял решение создать оцифрованный архив манускриптов Ватиканской библиотеки (10 лет, 40 млн. страниц, 45 петабайт) именно в FITS.
- FITS был предпочтен другим форматам (tiff, pdf, jpg, ...), поскольку он прост, открыт, хорошо документирован, «выживет» в процессе технической эволюции и будет «читабылен» в будущем.

Астрономия - единственная светская наука, в которую вовлечен Ватикан

# Наблюдения и модели

- Поставщики данных и сервисов внесли свой вклад в Виртуальную обсерваторию, предоставляя доступ к огромным коллекциям и архивам данных от радио- до гамма-диапазона длин волн.
- Совсем недавно стали появляться сервисы доступа к теоретическим моделям (например, коллекция теоретических спектров звёздных атмосфер в Испанской виртуальной обсерватории или синтетические модели звёздных популяций PEGASE во Французской виртуальной обсерватории, доступ к результатам космологических симуляций в Итальянской виртуальной обсерватории).

# МВО

- Таким образом, международная виртуальная обсерватория представляет собой мощную виртуальную среду, предназначенную для увеличения возможностей астрономических исследований и научного выхода данных. Виртуальная обсерватория интегрирует в единую систему гигантские астрономические архивы и базы данных, распределенные по всему миру, а также инструменты анализа данных и вычислительный сервис, используя при этом набор однородных стандартов и технологий.
- Международная виртуальная обсерватория включает в себя все значительные национальные и международные проекты по созданию виртуальных обсерваторий, основная цель которых – объединить существующие архивы наземных и космических инструментов и обеспечить исследователям и общественности удобный доступ к ним.

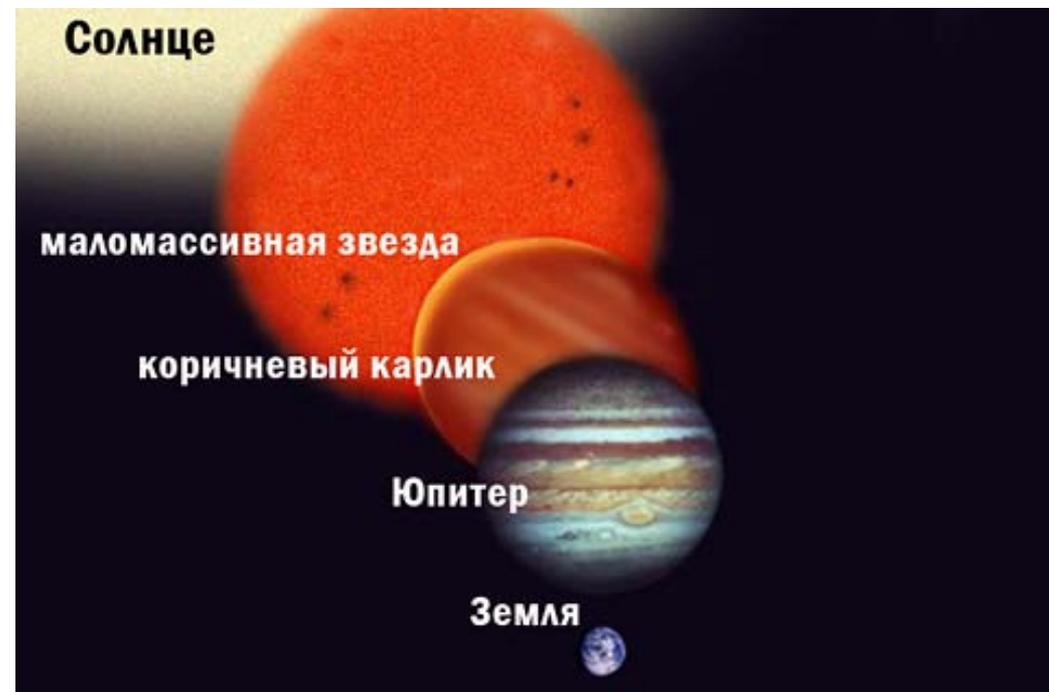
# МВО сегодня

- После нескольких лет интенсивного технологического развития ресурсы МВО достигли уровня зрелости, достаточного для их каждодневного использования в научных исследованиях. МВО находится на той стадии развития, когда астрономы начинают использовать её в повседневной научной работе.

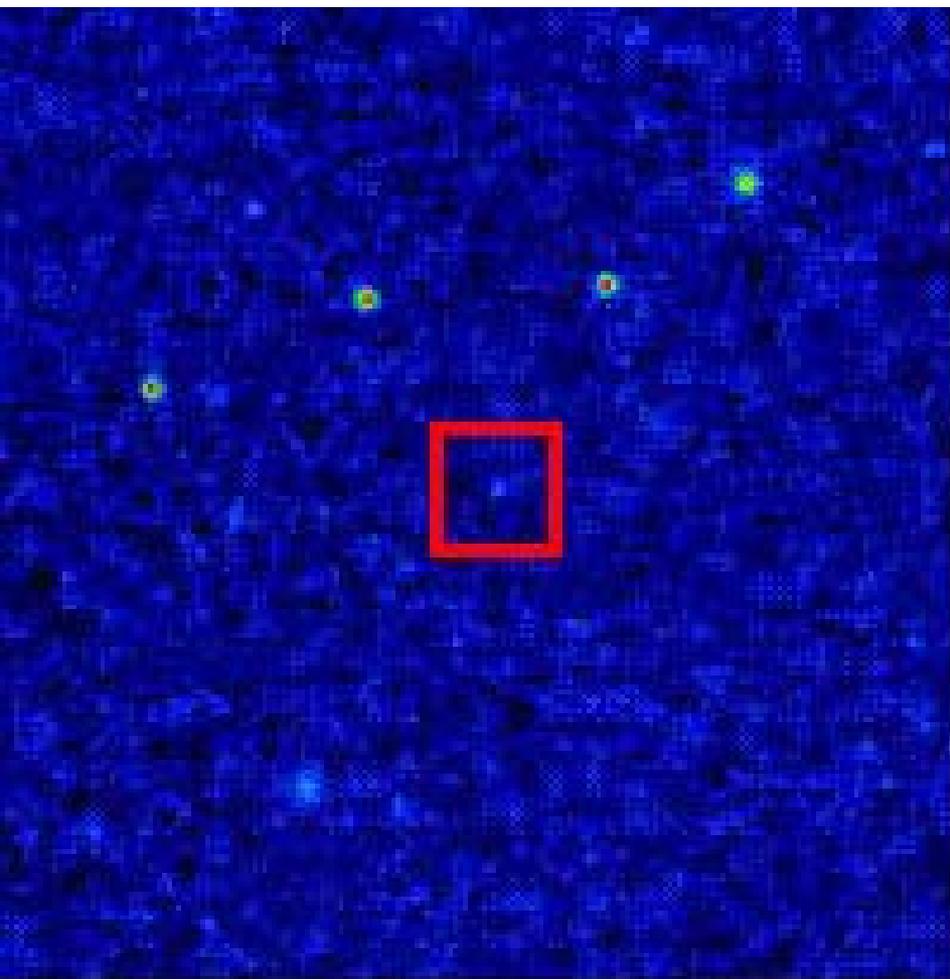
# Примеры задач звездной астрономии

- Коричневые карлики
- Квазары, наблюдаемые «с ребра»
- Рассеянные скопления
- Межзвездное поглощение
- Классификация затменных двойных
- База данных двойных звезд
- Собственные движения звезд
- Молодые звезды
- Яркие гиганты в близких галактиках

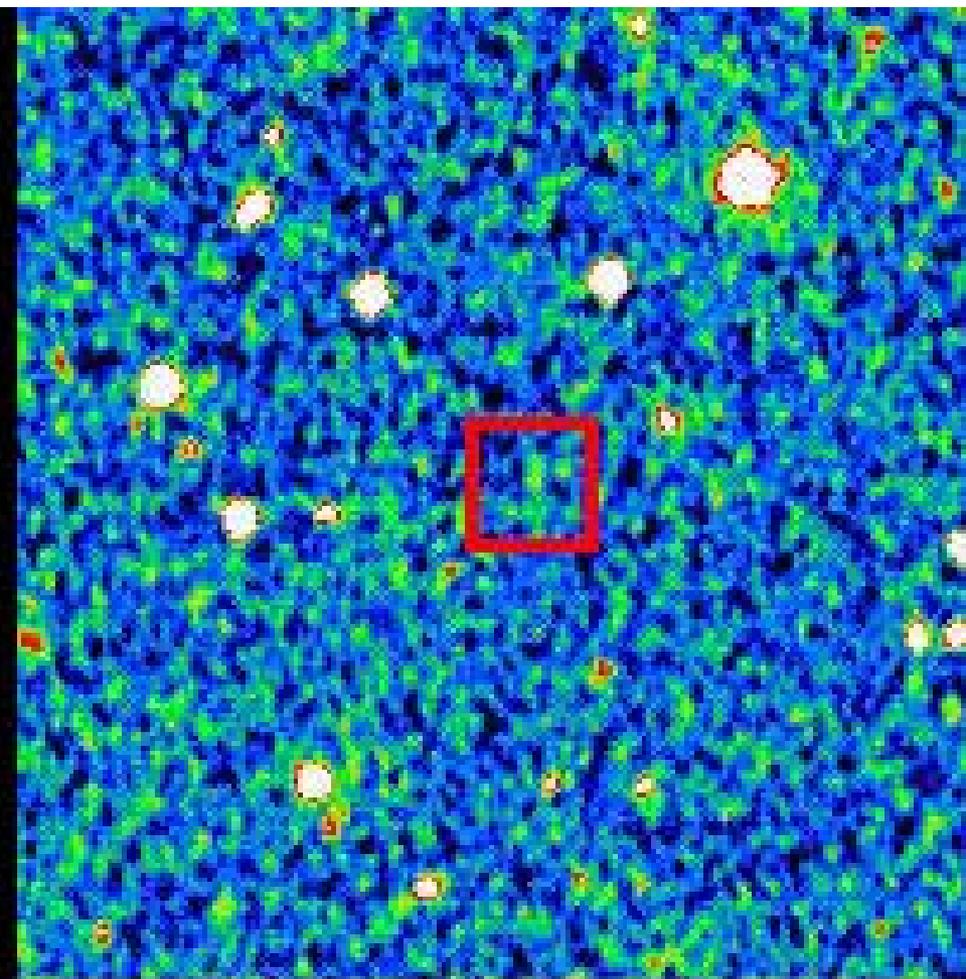
# Поиск кандидатов в коричневые карлики



- Показывают красные (i-z) цвета в SDSS и голубые (J-K<sub>s</sub>) цвета в 2MASS.
- Демонстрация прототипа NVO привела к открытию новых кандидатов в коричневые карлики. На поиск ушло две минуты (вместо недель или месяцев). Исследована 1/200 часть неба.
- Впоследствии результат был подтвержден спектроскопическими наблюдениями на крупнейшей наземной обсерватории Кек.

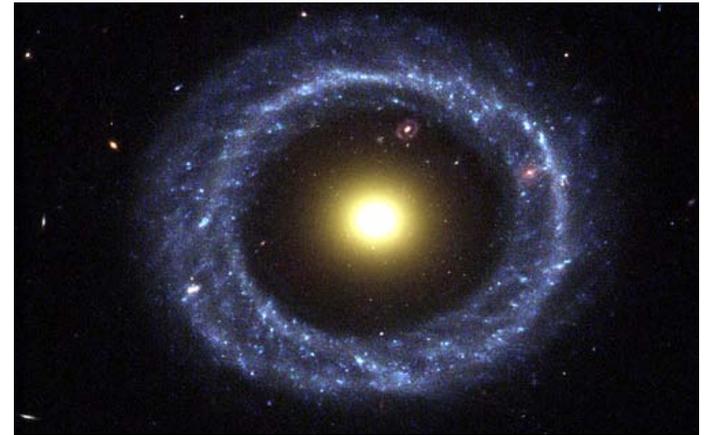


2MASS



SDSS

# Поиск квазаров второго типа



- Очень красные источники рентгеновского излучения
- Padovani et al. (AVO), комбинируя информацию с VLT, HST и Chandra, открыли 30 квазаров второго типа (ранее было известно 9 таких объектов)

# Рассеянные звездные скопления

В текущем состоянии галактической астрономии известно лишь 2% галактических рассеянных скоплений от их ожидаемого числа.

Информация о них накапливалась десятилетиями и даже столетиями в результате усилий множества исследователей.

Данные неоднородны и непригодны для систематического анализа.

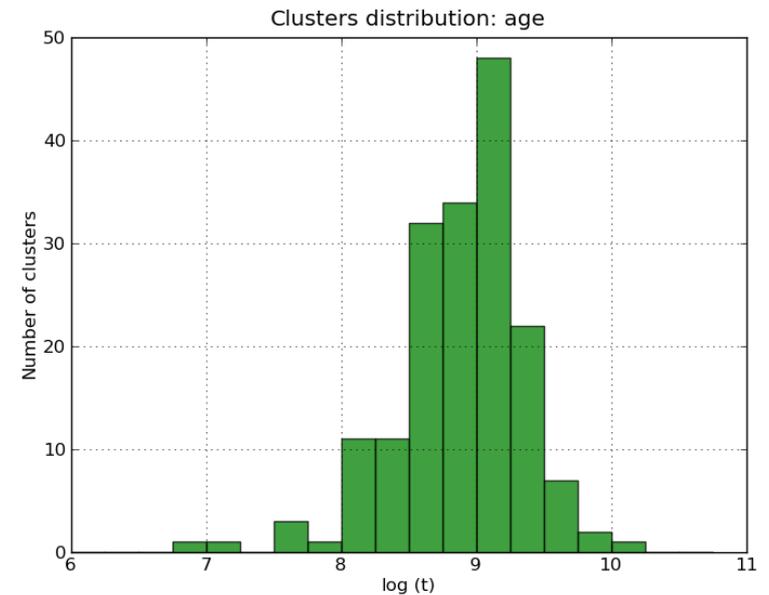
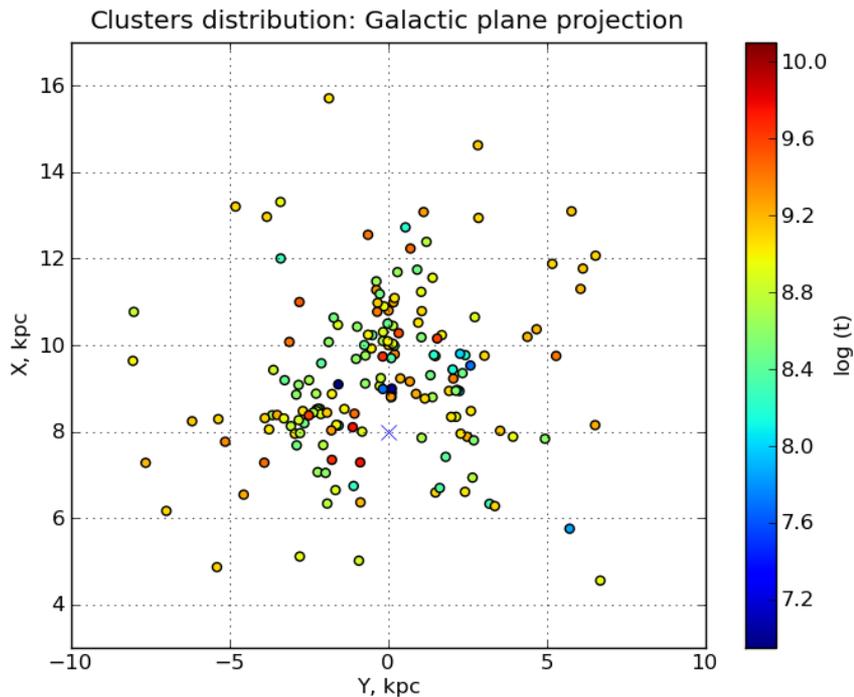
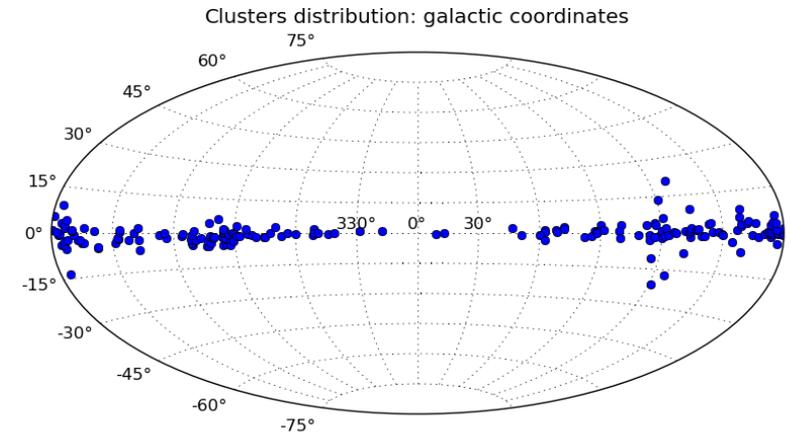
# Крупнейший каталог новых рассеянных звездных скоплений

- Kozlov, Glushkova, Zolotukhin, 2008, A&A, 486, 771; Glushkova et al., 2010, AstL, 36, 75
- Выполнен при помощи службы доступа к каталогам ГАИШ, SAI Catalog Access Service (Kozlov et al. 2007)
- 168 новых рассеянных скоплений, найденных и подтвержденных по многоцветной фотометрии в гигантских звездных каталогах (2MASS, UKIDSS)
- Представлены средства для получения новых научных результатов из данных каталога (+ данные BO) в режиме on-line
- Пример: в каталоге нет собственных движений, но за 5 минут их можно получить для скопления, а затем, например – кривую вращения Галактики.

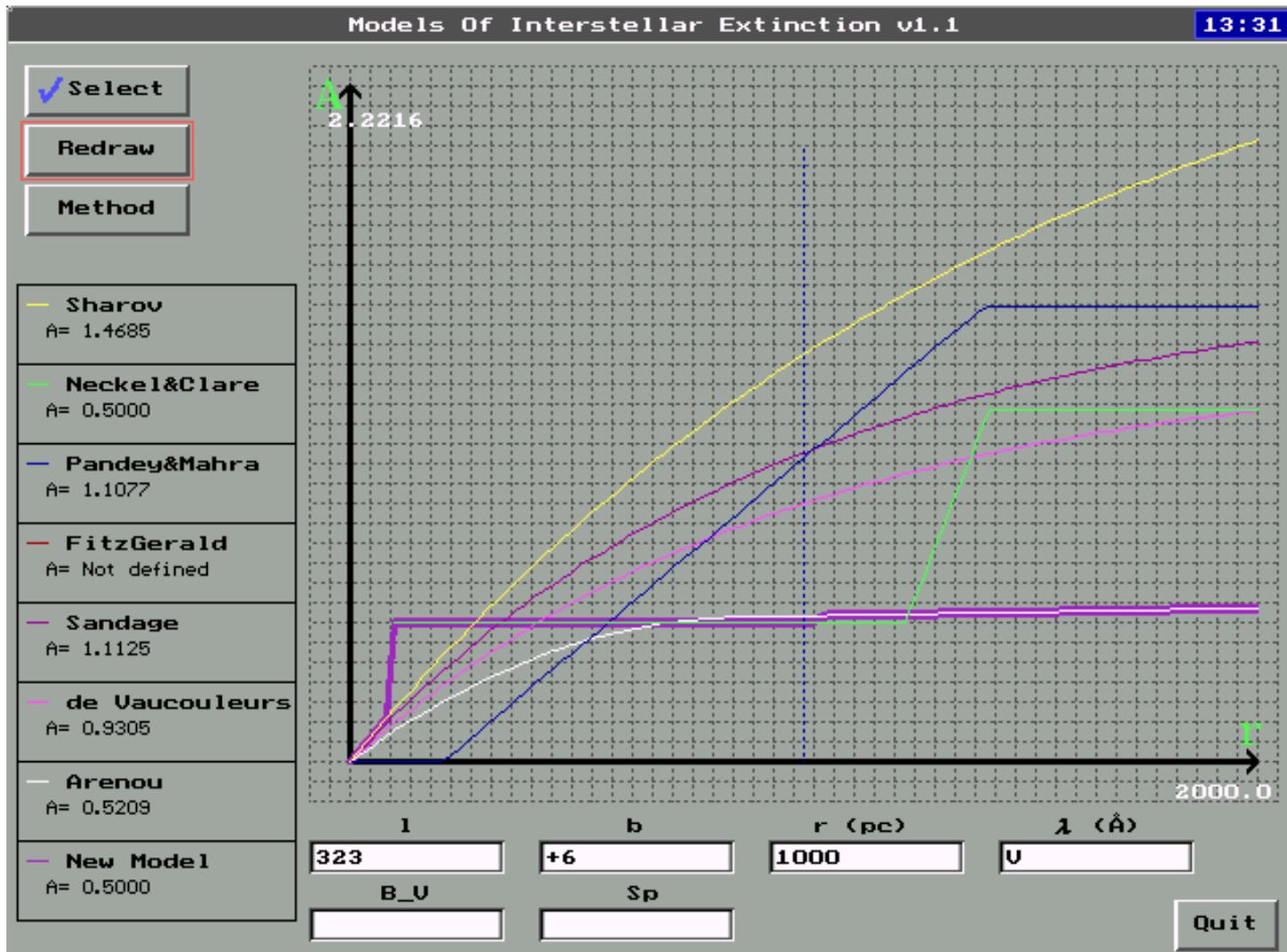
# Крупнейший каталог новых рассеянных звездных скоплений

- Зарекомендовавший себя метод весьма перспективен и для новых обзоров
- Можно ожидать десятки и сотни новых рассеянных скоплений в обзорах плоскости Галактики в ближнем ИК диапазоне – UKIDSS Galactic Plane Survey, обзоры телескопа VISTA.
- <http://ocl.sai.msu.ru>

# Крупнейший каталог новых рассеянных звездных скоплений



# Задача: построение карты межзвездного поглощения в Галактике



# Метод: использование многоцветной фотометрии звезд, содержащихся в больших обзорах. Этапы:

- Кросс-отождествление объектов
  - 2MASS, DENIS, SDSS, GALEX, UKIDSS, ...
  - 3-5 фотометрических полос
  - $10^7 - 10^9$  звезд
- Параметризация звезд
- Построение карты межзвездного поглощения, с учетом существующих (опубликованных) данных
- Пилотные результаты: Karpov & Malkov (2011, ASP Conf. Ser.), Malkov et al. (2011, ASS), Карпов и др. (2012, Астрофизический Бюллетень.), Malkov et al. (2012, Baltic Astronomy)

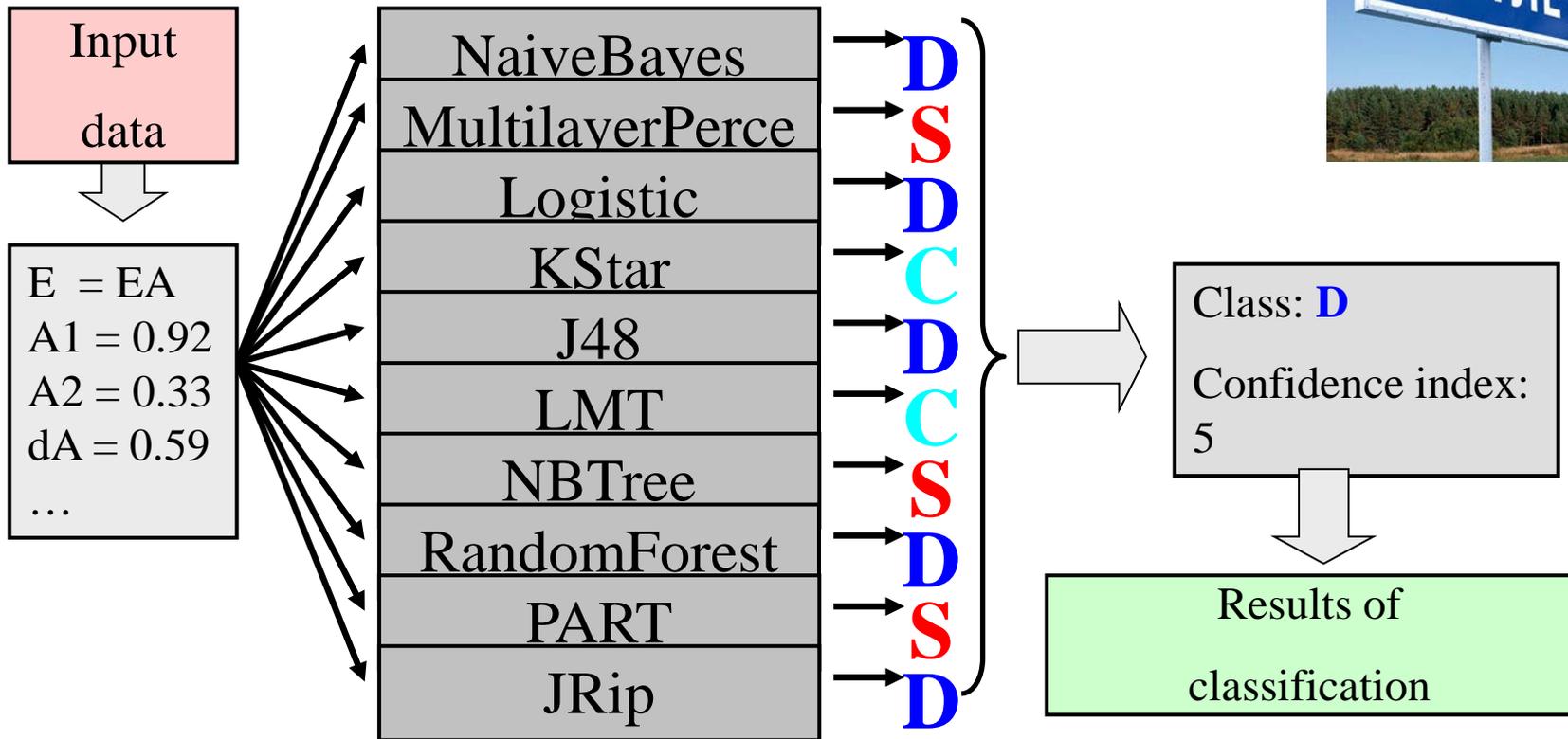
# Классификация и параметризация затменных двойных

- Создание каталога затменных двойных систем
- Разработка методики классификации затменных двойных на звездах с известным эволюционным классом
- Классификация затменных двойных
- *Параметризация затменных двойных*

## Classification scheme for eclipsing binaries

- D: Detached systems (0)
  - DM: Detached main sequence systems (155)
  - DR: Detached sub-giant systems (23)
  - DG: Detached giant or supergiant systems (22)
  - DW: Detached systems with white dwarf (14)
  - D2S: Detached symbiotic systems (5)
- S: Semi-detached systems (27)
  - SA: Classical Algols (374)
  - SC: Cool semi-detached systems (4)
  - SH: Hot semi-detached systems (33)
  - S2: Late stage semi-detached systems (0)
    - S2C: Cataclysmic systems (32)
    - S2H: High-mass X-ray binaries (3)
    - S2L: Low-mass X-ray binaries (3)
- C: Contact systems (126)
  - CB: Near-contact systems (104)
    - CBF: Near-contact F systems – primary is at Roche lobe (8)
    - CBV: Near-contact V systems – secondary is at Roche lobe (10)
  - CE: Early-type contact systems (17)
  - CW: Late-type contact systems (1)
    - CWA: Late-type contact A systems – primary is larger (99)
    - CWW: Late-type contact W systems – secondary is larger (118)
  - CG: Giant contact systems (4)

Классификация в инфраструктуре РВО базируется на системе AstroGrid (UK), установленной в Суперкомпьютерном центре РАН и в ИПИ РАН



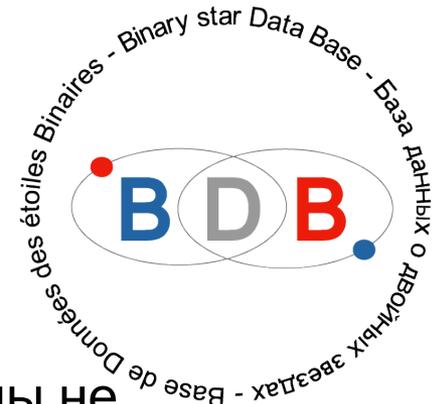
### Bibliography:

- Malkov et al. 2010, MNRAS 401, 695
- Malkov 2007, MNRAS 382, 1073
- Malkov et al. 2007, A&A 465, 549
- Malkov et al. 2006, A&A 446, 785

# База данных двойных звезд

- Существуют базы данных по различным типам двойных звезд (WDS, ОКПЗ, CBS), но не существует базы данных, синтезирующей все типы двойных. BDB призвана закрыть этот пробел.
- Цель создания BDB – предоставить пользователю каталогизированные данные о двойных
  - интегрируя в BDB каталоги всех типов двойных (CCDM, ORB6, ...), а также каталоги астрофизических параметров – масса, радиус, температура – (Свечников, Budding, ...),
  - предоставляя доступ к базам данных о двойных системах (ОКПЗ, SB9, ...) и базам данных общего назначения (SIMBAD, ADS, ...)

# База данных двойных звезд



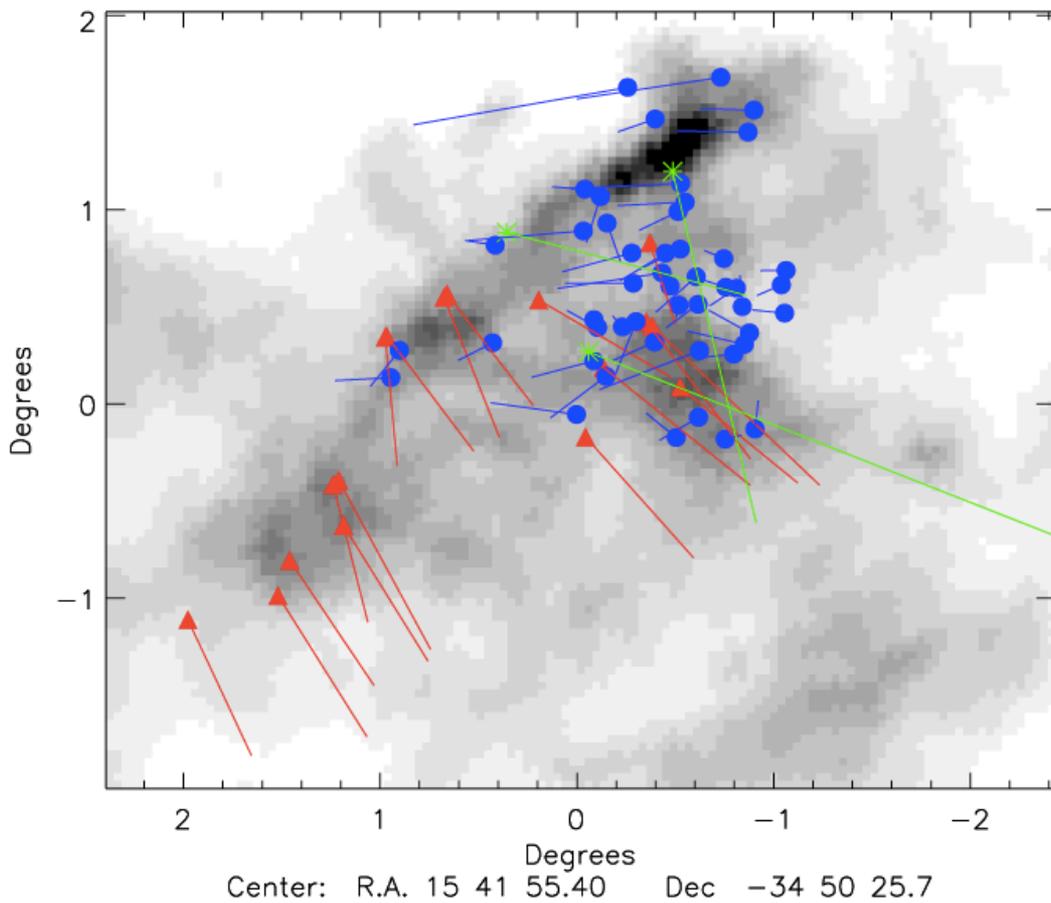
- Основные проблемы:
  - решение проблем обозначений (единой системы не существует) и кросс-идентификации компонентов в разных каталогах
  - создание и поддержка списка ассоциированных каталогов / баз данных двойных звезд (прежде всего российских: ОКПЗ, ТДЗ, каталоги Свечникова и пр..).
  - остальное сделают инструменты Виртуальной обсерватории.
- Научные приложения:
  - создание каталогов астрофизических параметров звезд,
  - получение фундаментальных эмпирических зависимостей,
  - уточнение сценариев эволюции двойных звезд, ...
- [bdb.inasan.ru](http://bdb.inasan.ru)

# Собственные движения объектов из темных облаков в созвездии Волка

- Lopez Marti, Jimenez-Esteban, Solano, 2011, A&A, 529, 108
- В последние годы по данным телескопа Spitzer открыто множество звездных и дозвездных объектов – кандидатов в объекты из комплекса звездообразования в Волке
- Авторы занимались кинематическим подтверждением принадлежности кандидатов к комплексу звездообразования по данным Виртуальной Обсерватории (кросс-матчинг с каталогами собственных движений)
- Выделено две группы объектов: одна с существенными коррелированными собственными движениями, согласующимися с собственными движениями других популяций пояса Гулда (что указывает на предполагаемую связь между поясом Гулда и комплексом в Волке), и вторая группа с незначительными случайными с.д., представляющая фоновые объекты



# Собственные движения объектов из темных облаков в созвездии Волка



# Молодые звезды и коричневые карлики вокруг Алнилама и Минтаки

- Caballero, Solano, 2008, A&A, 485, 931
- Авторы исследовали 2 области радиусом 45' вокруг молодых звезд из пояса Ориона
- Виртуальная Обсерватория: 2MASS, DENIS, Tycho-2, а также рентгеновские, инфракрасные и спектральные архивы
- По избытку ИК-излучения, наличию лития в абсорбции и раннему спектральному типу обнаружено 136 экстремально молодых объекта, 2 молодых коричневых карлика и 289 других кандидатов в члены ассоциаций
- Вывод: оба региона похожи на скопление вокруг  $\sigma$  Ori, но немного старше, протяженнее и обладают меньшей радиальной концентрацией

1940 публикаций в ADS на октябрь 2013 г.  
упоминают Virtual Observatory в аннотации

≈ 220 из них являются полноценными  
астрономическими исследованиями,  
опубликованными в международных рецензируемых  
журналах

(see <http://ivoa.net/newsletter/006/>)

Первые диссертации по результатам исследований в  
помощью инструментов ВО были защищены в России в  
2009 г. (кандидатская, И.Золотухин) и 2010 г. (докторская,  
И.Чилингарян)

# Актуальные проблемы ВО

- Существенное развитие инструментов наблюдения для получения данных о звёздах, открытость данных и сплочение исследователей в сообщества диктуют новые требования к инфраструктуре исследований в астрономии. Виртуальная обсерватория уже имеет свои плоды, но для решения предстоящих проблем она должна развиваться в определённых направлениях.

# Неоднородность данных

- Проблемой, которая несёт за собой много последствий, является неоднородность накопленных до сих пор данных. Каждый обзор неба и каждый из множества каталогов создавались разными людьми или исследовательскими группами для решения определённых задач, поэтому состав и структура данных в каталогах бывают ограничены теми целями, которые преследовались их создателями. Каталоги имеют неоднородную структуру и различную семантику полей и значений, множество важных, но никак не формализованных комментариев. Совместная обработка данных на стыке каталогов становится не автоматизируемой и требует программирования.
- Использование в Виртуальной обсерватории каталогов в их первоизданном виде оправдывается целью сохранения уникальных и не теряющих актуальности наблюдений в астрономии, однако в результате приходится решать одни и те же проблемы неоднородности множество раз от задачи к задаче. Неоднородность данных является первейшей проблемой, отнимающей время и ресурсы при разработке научных приложений.

# Кросс-идентификация

- Работа по сопоставлению каталогов в рамках виртуальной обсерватории ведется на уровне создания таблиц кросс-идентификации. Примерами такого подхода могут служить популярная база астрономических данных Simbad (<http://simbad.u-strasbg.fr>) и ее прототипы. Также создаются системы с ограниченными наборами интегрированных каталогов. И в том, и в другом случае установление взаимосвязи является кропотливой работой. Различия в семантике данных порой столь неочевидны, что могут выявляться со временем при обнаружении противоречивости результатов, появлении неверных связей.

# Предметные области

- С возрастанием количества каталогов, исследуемых в них параметров, разновидностей решаемых задач и интенсивности использования данных различного происхождения при их решении назревает необходимость в семантических подходах к работе с каталогами и к решению научных задач. Исследовательские группы должны взаимодействовать в рамках сообществ с целью разработки спецификаций предметных областей, в которых они работают. Разрабатываются онтологии предметных областей, стандартизируются концептуальные схемы для представления данных в приложениях сначала по общим разделам знаний, используемым практически во всей астрономии, затем более специальным.
- В рамках Международной виртуальной обсерватории наработки в области семантического моделирования предметных областей присутствуют, в частности онтология астрономических объектов, ряд схем (называемых моделями данных) для представления разного рода данных, например, PhotDM для представления фотометрических данных, VO Event для описания физических явлений и процессов и другие

# Методы анализа данных

- Ещё одной стороной, требующей существенного развития в Виртуальной обсерватории, является накопление общедоступных реализаций методов вычислений и анализа данных, характерных как для астрономии в целом, так и в конкретных областях исследования. Помимо исследуемых данных астрономы нуждаются в средствах их анализа.
- Необходимо создавать коллекции универсальных настраиваемых сервисов данных, независимых от источников данных, реализующих методы анализа. Общедоступными и легко настраиваемыми на любые данные, на пересечение различных массивов данных и параметров должны быть средства статистического и визуального анализа, в частности, создания диаграмм. Необходимы мощные средства редукции данных, поиска похожих объектов, поиска отличающихся объектов (ошибочных или необычных) и решения других универсальных задач. Необходимы более узконаправленные средства, связанные с конкретными методами, законами в астрономии.

# Эффективное представление данных

- В Виртуальной обсерватории ожидается существенное развитие технологий, связанных со скоростью обработки данных при анализе. Исходные каталоги обычно используют файловое хранение с записями фиксированной длины (это делается традиционно для удобства обработки подручными средствами, и такая структура привычна для астрономов). В таком представлении нет индексации, нет предварительного агрегирования больших массивов данных, хранение данных не оптимизировано. Результатом этого является медленная работа обрабатывающих их приложений. Задача эффективного представления данных требует решения.

# Проблема ресурсов

- Используемые подходы к решению задач часто требуют физического перемещения внушительных массивов данных через интернет или на носителях. При решении некоторых задач требуется множественный прогон алгоритмов и методов. Некоторые каталоги интенсивно пополняются данными от версии к версии, ожидается появление обзоров колоссальных объёмов.
- В Виртуальной обсерватории необходимо использовать при автоматической обработке и предоставлять в пользование исследователям средства распределённого хранения и параллельной обработки на основе технологий больших данных и облачных технологий.

# Заключение

- Средства Виртуальной обсерватории в области звёздной астрономии как науки с интенсивным использованием данных стали её неотъемлемой частью и двигателем развития. Об этом говорит их использование при решении множества научных задач в этой области. В приведённых примерах задач, решённых в среде Виртуальной обсерватории, использовались совместно данные из разных астрономических каталогов.
- Открытость и обилие научных данных наблюдения, колоссальный рост объёмов данных и их обработки предъявляют новые требования к данным и к Виртуальной обсерватории, являющейся их источником для научных исследований.

# Благодарности

- И. Золотухину за помощь в подготовке материала.
- РФФИ (гранты 12-02-31904, 12-07-00528), Федеральное агентство по науке и инновациям (грант 02.740.11.0247), Программе Президиума РАН Поддержка ведущих научных школ (грант НШ-3602.2012.2), ФЦП «Научные и научно-педагогические кадры инновационной России» на 2009 - 2013 годы.
- Организаторам конференции – за приглашение.
- Аудитории – за внимание.