

Корпоративная аналитическая база данных статей: принципы организации и методика формирования.

В предлагаемых вашему вниманию следующих трех докладах мы хотели бы рассказать о своем опыте объединения усилий нескольких десятков библиотек с целью создания корпоративной базы данных, содержащей библиографические описания журнальных статей, сопровождаемых краткими аннотациями. Мы расскажем о том, как организована совместная работа, каким образом формируется корпоративная база данных, каковы первые итоги работы, проблемы и перспективы нашего проекта.

1. Почему мы вместе

Прежде всего, несколько слов о том, почему мы вместе.

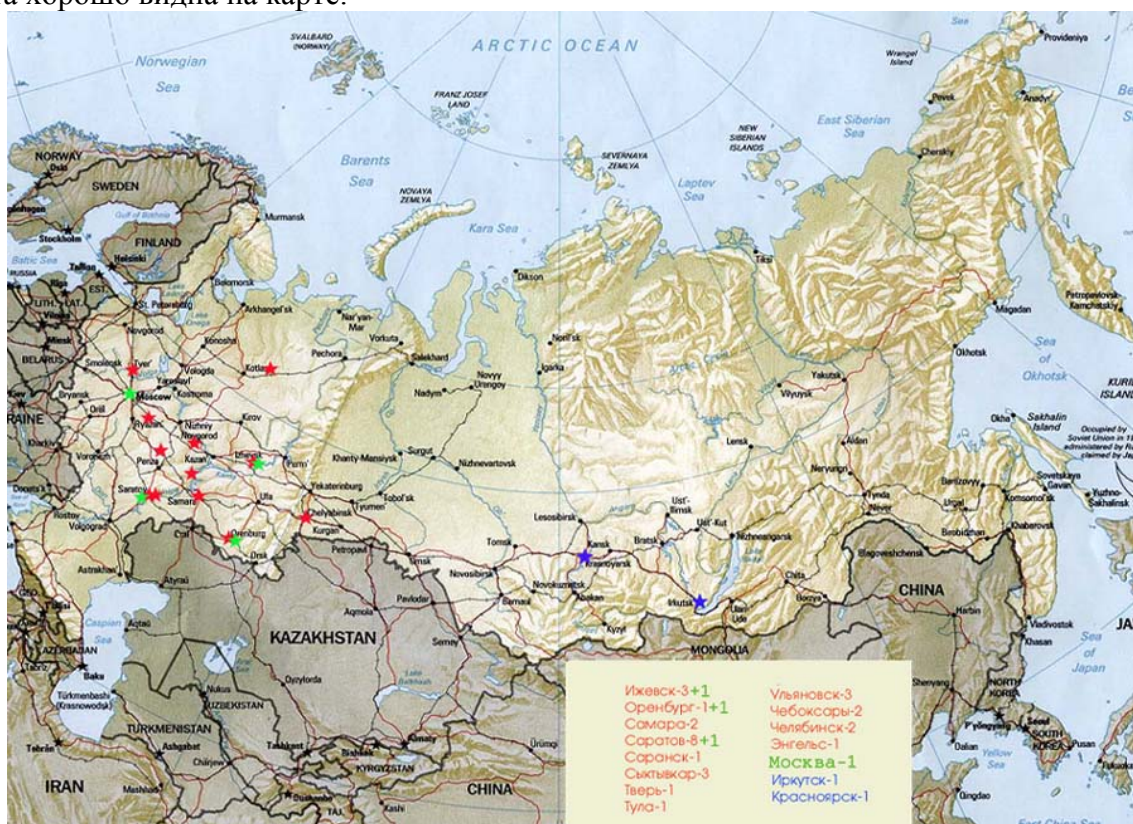
Хорошо известно, что в большинстве библиотек большим читательским спросом пользуются картотеки и базы данных, содержащие так называемую роспись журнальных статей. Длительное время они формировались на основе комплекта карточек Книжной Палаты. Однако в последние несколько лет многие библиотеки были вынуждены начать самостоятельную роспись журналов. В конечном итоге сложилась ситуация, когда в библиотеках, с одной стороны, многократно расписываются одни и те же издания, а, с другой стороны, библиотеки, особенно не очень крупные, физически не могут расписать необходимое число журналов для полноценного обслуживания пользователей.

Для того чтобы устранить такое дублирование в работе библиотек, и было создано наше объединение. Идея проекта возникла в ходе работы прошлогодней конференции LIBCOM-2000, когда представители 4 библиотек - Саратовского, Челябинского и Удмуртского государственных университетов и Самарского аэрокосмического университета - собрались для обсуждения своих проблем в работе. Они оказались общими и понятными для всех. Мы рассказали друг другу как работаем в своих библиотеках, какие трудности испытываем. Тогда же мы договорились об общих принципах и методах совместной работы.

В результате объединения получился проект, который сегодня представляет собой нечто большее, чем простое присоединение друг к другу. Наша совместная работа строится по простому принципу: каждый участник расписывает от 5 до 10 журналов, а в обмен получает все, что расписывают другие участники проекта.

2. Состав участников

Сейчас в проекте принимают участие 28 библиотек из 12 городов РФ, 4 библиотеки проходят процедуру присоединения к проекту, 2 решают вопрос о присоединении (рассматривают документы и образцы библиографических записей) и 1 выходит из проекта. "География" проекта хорошо видна на карте.



Среди библиотек-участниц 21 вузовская библиотека (в том числе 11 библиотек классических университетов), 5 национальных и областных библиотек, 1 областная библиотека для детей и юношества, 2 центральные городские библиотеки.



Библиотеки самые разные по составу. Есть библиотеки, в которых в работе задействованы отделы периодических изданий, справочно-библиографические отделы и отделы автоматиза-

ции, есть и такие библиотеки, штат которых насчитывает всего 5-6 человек и всю работу по проекту ведет один из них.

Этими силами производится полная аналитическая роспись 238 названий периодических журналов научной и научно-популярной тематики.

3. Организация обмена информацией.

При выборе способа обмена информацией мы исходили из того, что необходимо обеспечить устойчивый обмен данными с использованием минимальных технических средств. Наиболее подходящими в этом смысле, на наш взгляд, являются широко распространенные листы рассылки, когда для передачи и получения сообщений библиотеке необходимо лишь иметь электронную почту. Такая возможность в настоящее время есть в большинстве библиотек.

4. Организация работы и принятие общих решений.

Для координации работ по проекту на очередной календарный год из числа участников проекта избирается библиотека-координатор. Все вопросы, требующие совместного решения, выносятся координатором на голосование через лист рассылки. При этом указывается вопрос и сроки голосования. В ходе голосования может быть проведено обсуждение (опять же через лист рассылки). Все решения принимаются простым большинством голосов и являются обязательными для всех участников проекта.

Ежеквартально библиотека-координатор подводит итоги работы и передает в лист рассылки сводный список журналов, роспись которых поступила в лист. В это же время ликвидируются "пропуски" и передаются напоминания тем библиотекам, которые не выдерживают оговоренных сроков предоставления росписей.

5. Регламентирующие документы

Работа по проекту регламентируется Договором, который виртуально подписывается всеми участниками.

В Договоре определены права и обязанности участников, их ответственность, оговорена процедура приема новых участников. Каждый участник проекта обязуется расписать "свои" журналы в течение семи рабочих дней с момента их поступления в библиотеку, для исправления ошибок установлен срок в три рабочих дня.

Для вступления в проект все желающие знакомятся с пакетом общих нормативных документов. Желающим "кандидатам" предлагается выбрать из своего списка получаемых периодических изданий журналы, которые бы им удобно было расписывать. Этот список выносится на всеобщее обсуждение и после общего голосования он за библиотекой - "кандидатом" закрепляется. На время обсуждения списка журналов библиотеке - "кандидату" открывается доступ к списку рассылки, чтобы они могли в режиме реального времени ознакомиться с технологией совместной работы, получить реальные росписи журналов, проверить их качество на своих программных продуктах.

После принятия решения о вступлении в проект библиотека - "кандидат" переходит в разряд "новичков". Она должна пройти у нас своеобразное "испытание": каждый из новых участников проекта должен предоставить на общий суд свою работу - подготовить росписи закрепленных изданий по нашей методике, выслать их в коллективный список. После получения от "новичка" полного комплекта росписей закрепленных за библиотекой журналов, она получает от координаторов всю базу данных за год.

Т.к. проект изначально осуществляется на безфинансовой основе, единственным "наказанием" для участников может быть исключение из проекта.

В Договоре прописана также процедура выхода из проекта.

В Приложениях к Договору приводятся списки расписываемых журналов и закрепленных за ними библиотек, правила заполнения полей, порядок приема и передачи данных, принятый в рамках проекта рабочий список рубрик и подрубрик.

О списке рубрик и подрубрик хочется остановиться особо.

В нашем проекте, как уже говорилось, работают библиотеки разных типов и ведомств. В каждой из них сложился определенный стиль работы, традиции. Для систематизации используются различные системы классификаций: УДК, ББК, ГРНТИ. Для того, чтобы достичь некоторого компромисса, при формировании рубрикатора для совместной работы было принято следующее решение. Чтобы в максимальной степени использовать опыт работы конкретной библиотеки с классификационными системами УДК или ББК и не обязывать библиотеки переходить с одной системы на другую, был сформирован и согласован свой список рубрик и подрубрик.

При этом согласовывались лишь текстовые формулировки. Сами формулировки брались из таблиц ББК или УДК, в зависимости от того, в какой из систем, на наш взгляд, формулировки для данной предметной области в большей степени подходят для нашей работы. Вводились и новые понятия. К каждой рубрике и подрубрике рубрикатора затем приписывался индекс ББК или УДК. В дальнейшем планируется проставить индексы Дьюи и ГРНТИ.

Конкретный раздел рубрикатора в настоящее время закрепляется за "профильной" библиотекой, которая следит за развитием своей предметной области и оперативно вносит предложения по совершенствованию рубрикатора.

К таким решениям мы пришли после попыток использовать "независимые" системы - систему Дьюи и Авторитетный файл предметных рубрик РНБ. От первой мы отказались, т.к. специалисты библиотек не были готовы к ее использованию. Со второй возникли чисто технические проблемы ее использования в наших программных продуктах. Эти проблемы в настоящее время решаются. К вопросу об использовании рубрик РНБ мы вернемся к началу следующего года.

5. Используемое программное обеспечение

Для формирования корпоративной базы данных участниками проекта используется различное программное обеспечение.



18 библиотек участниц готовят росписи журналов с помощью различных версий программы "Библиотека", разработанной в МГУ. Это в основном, сетевой вариант 4 версии различных модификаций. 11 библиотек используют различные версии АИБС "МАРК". В процентном соотношении используемое программное обеспечение представлено на диаграмме 2. Особенностью этих программных средств является то, что оба продукта формируют библиографические записи в базах данных в формате US MARC. А поскольку большинство участников использует пакет "Библиотека", то в качестве обменного формата данных был взят формат записей этой программы.

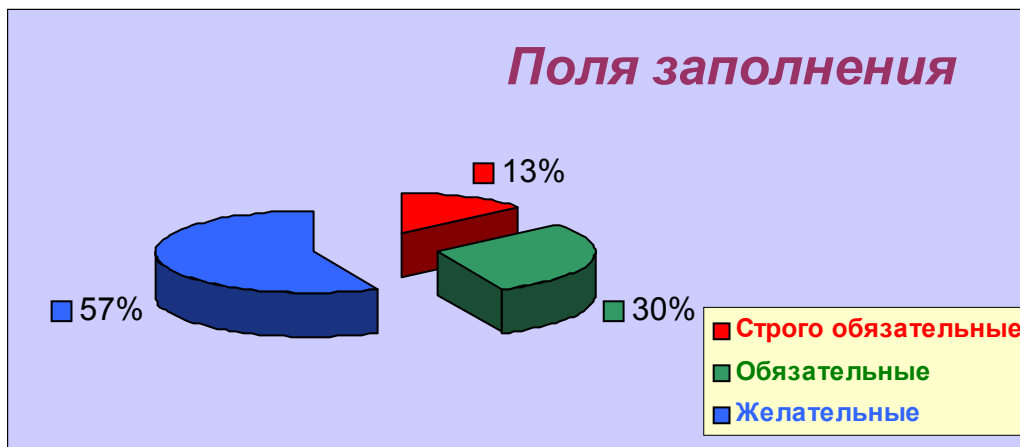
Реализация формата US MARC в АИБС "Библиотека" и АИБС "МАРК" немного отличаются друг от друга. Поэтому для библиотек, использующих АИБС "МАРК" был разработан и безвозмездно передан всем участникам небольшой конвертор.

6. Методика создания корпоративной базы данных.

Во время предварительного этапа работы библиотеки – инициаторы проекта провели сверку форматов записей обоих программных продуктов и анализ перечня полей, необходимых для аналитической росписи периодических изданий. Был составлен список кодов полей формата US MARC для этого перечня. В результате была составлена и представлена всем участникам проекта таблица соответствия заполняемых полей в обеих задачах (представить таблицу) и выделены три группы полей, используемых для библиографического описания периодического издания.

- **Поля, строго обязательные для заполнения.** Без их заполнения запись о расписываемой статье считается ошибочной.
- **Поля, обязательные для заполнения.** Это поля, которые обязательно надо заполнять, если в расписываемых статьях есть соответствующая информация.
- **Поля, желательные для заполнения.** Это поля, которые заполнять не обязательно, но если информация вносится, то она должна вноситься именно в указанные поля.

Общее количество полей и подполей US MARC, на основе которых создается сводная база данных, - 90. Из них **полей, строго обязательных для заполнения - 12, обязательных для заполнения полей - 27. Полей, желательных для заполнения - 51.**



При составлении перечня полей с учетом дальнейшего развития проекта были введены поля, которые отсутствовали в стандартных настройках полей обеих задач. Например, было принято решение о введении в список желательных для заполнения поля 080a – индекс Дьюи и 856u – адрес URL источника.

Все участники проекта, вместе с перечнем полей получила краткие комментарии по их заполнению. Эти комментарии были составлены на основе методических инструкций и практического опыта работы библиотек-участниц проекта.

В кратких комментариях к списку заполняемых полей, перечислены правила и особенности заполнения полей, регламентируются общие договоренности между библиотеками-участницами, фиксируются решения, принимаемые на основе общего голосования по каким-либо спорным вопросам.

Одним из основных условий создания корпоративной базы данных является **полная аналитическая роспись** издания.

Полнота расписываемого издания проверяется согласно его содержанию. Объем расписываемого издания регламентируется специально введенным полем.

Заполнение поля 520a – «Аннотация» основное условие аналитической росписи в совместной работе. Аннотация призвана дать читателю краткую, точную и полную информацию о содержании статьи.

Поля из групп **обязательно заполняемых полей** заполняются согласно нормативным документам, предоставляемых координаторами проекта. Нормативные документы уточняются в начале каждого квартала работы и получаются всеми участникам через список рассылки.

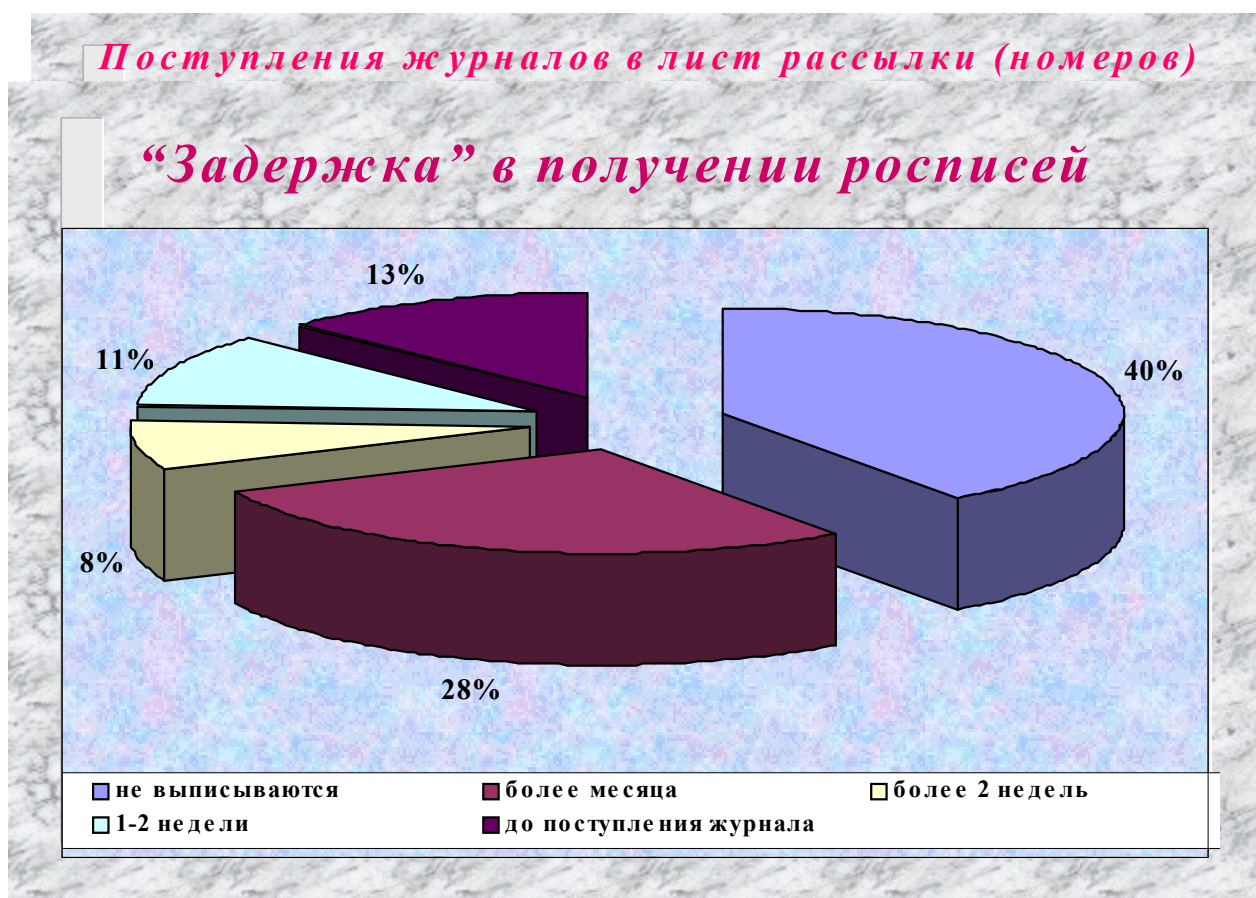
В каждой библиотеке по-разному подходят к их использованию. Там, где есть отделы автоматизации, из них подготавливают машиночитаемые словари и подключают к соответствующим полям для заполнения информации. Это очень ускоряет ввод данных и уменьшает

вероятность ошибок при ручном вводе. Там, где таких отделов нет, эти документы служат важным источником для ввода и проверки заполняемых данных.

7. Немного статистики.

Всего в рамках проекта за 10 месяцев работы в сводную базу данных поступило более 20 тыс. библиографических записей, представляющих роспись более тысячи номеров журналов.

На приведенном рисунке представлен график поступлений. Видно, что нам не удалось избежать спада поступлений, вызванного периодом летних отпусков, однако летние "долги" уже сейчас активно ликвидируются.



Большинство участников соблюдает оговоренные сроки выполнения росписей, хотя есть и задержки, связанные как с объективными, так и с субъективными причинами.

На следующем рисунке показано поступление журналов в октябре месяце в Научную библиотеку Саратовского государственного университета. Всего в лист рассылки поступило 154 номера. Из них 63 номера из журналов, которые библиотека не выписывает. 45 номеров (28%) - поступили с задержкой более месяца, 12 (8%) с задержкой более 2 недель, 18 номеров (11%) пришли в библиотеку в течение 1-2 недель, 20 номеров (13%) - досрочно. Из журналов, которые библиотека выписывает, а это 95 названий из 154, 38 номеров, т.е. 40% приходят в пределах 2 недель.

8. Каков же результат нашей совместной работы.

Состояние базы данных на 8 ноября в библиотеке Челябинского государственного университета представлено ниже.

Из них видно, что объем данных в базе составляет 20 436 записей. Это занимает более 18 Мб дисковой памяти. 66% информации заносится руками, что составляет более 12 Мб информации. Средний объем записи - 897 байт. Количество полей, содержащееся в каждой записи, 54 из 90 регламентированных согласно принятым документам.

Не все записи удовлетворяют правилам заполнения -2% из них, например, не имеют заглавия источника. 1% записей не имеет строго обязательного поля - кода записи.

Практически все записи содержат предметные рубрики и ключевые слова. 87% записей - с аннотациями, длина которых - более 220 байт, т.е. - четверть записи.

Корпоративная аналитическая база данных (8.11.01, НВ ЧелГУ)			
Общая статистика:	Всего записей в базе	-	20 471
	Просканированно записей	-	20 446
Длина всех записей в US MARC - 18.344 Мб.			
Величина заполненных данных - 12.18 Мб (66.4% MARC-записи)			
Средняя длина записи - 897.2 б			
Среднее количество полей в записи - 53.73			
Статистика заполнения полей			
	(байт)	Всего полей:	в %
Средняя длина названия издания	- 19.79	20087	98
Средняя длина таблицы для авторов	- 11.75	22865	100
Средняя длина таблицы для заглавия	- 49.46	20431	100
Средняя длина таблицы для кол. автора	- 29.98	3006	14
Средняя длина таблицы для предм. рубрик	- 14.84	52545	100
Средняя длина таблицы для кл. слов	- 15.20	116434	100
Средняя длина таблицы для кода записи	- 9.15	20333	99
Средняя длина таблицы для адреса URL	- 19.17	9188	45
Средняя длина таблицы для УДК, ББК, Дьюи	- 5.30	29000	75
Средняя длина аннотации	- 227.70	17867	87
	Всего:	В среднем в одной записи:	
Кол-во строго обязательных полей	- 230946	94.2%	(11.30 из 12)
Кол-во обязательных полей	- 135429	24.7%	(6.67 из 27)
Кол-во желательных полей	- 109433	10.5%	(5.35 из 51)
Среднее кол-во предметных рубрик в записи - 2.58			
Среднее кол-во ключевых слов в записи - 5.70			
Занесены индексы УДК, ББК, Дьюи в 74.72 % записях			
Занесены адреса URL в 44.96 % записях			

Интересны цифры по статистике так называемых информационно-содержательных полей.

В каждой записи в среднем по 2.5 рубрик. Каждая запись сопровождается более, чем 5 ключевыми словами. Почти 75% базы данных – систематизировано с помощью классификационных систем, т.е. в записях имеются индексы ББК, УДК, Дьюи.

Интересны показатели в последней строчке. Но это - тема особого выступления. Пока можно лишь сказать, что 45% наших росписей предварительно готовы для создания системы электронной доставки документа.

Мы попытались проанализировать тенденцию нашей совместной работы. Следующая таблица показывает качественное изменение записей корпоративной базы данных с течением времени.

<i>Сравниваемые параметры (в среднем в записи)</i>	<i>I кв.</i>	<i>II кв.</i>	<i>III кв.</i>
Количество полей	48,02	51,14	51,73
Длина записи (байт)	817,55	868,06	903,03
Длина аннотации (байт)	212,26	224,33	244,76
Количество предметных рубрик	2,27	2,54	2,72
Количество ключевых слов	5,11	5,51	6,03

8. Некоторые итоги нашей совместной работы

В целом, проект за 10 мес. своего рабочего существования показал свою жизнеспособность. Если в начале работ в проекте изъявили желание участвовать около 12 библиотек, то в феврале его участниками были уже 26 библиотек. Сейчас в нашей команде - 29, 4 - находятся в стадии присоединения, так называемые "кандидаты".

Единственным условием участия в проекте (помимо желания) является соблюдение общих договоренностей, которые, напомним, принимаются всеми участниками простым большинством голосов. У нас, как и во всяком - реально действующем проекте, еще остается много вопросов. Они касаются методики формирования корпоративной базы данных, контроля над выполнением принятых решений и по другим вопросам.

Но уже сейчас ясно, что мы научились слушать и слышать друг друга, договариваться, принимать общие решения. За это время многие библиотеки получили ценные профессиональные консультации. Отработан механизм взаимодействия между разно- ведомственными библиотеками.

Объем корпоративной базы данных составляет почти 20 тысяч библиографических записей. Вряд ли какая-нибудь из библиотек-участниц смогла бы за это время сформировать такую базу данных своими силами. Наша общая работа необходима всем нам.

Уже в ходе подписки на первое полугодие 2002 года некоторые библиотеки, используя материал корпоративной базы данных, внесли коррективы в репертуар выписываемых изданий.

Главное, мы поняли, что можем работать вместе и можем и дальше развивать этот проект.

А в планах у нас увеличение числа расписываемых региональных изданий, создание сайта проекта, формирование с использованием отработанной модели взаимодействия корпоративной базы данных "серой" литературы. Уже проведены пробные работы по электронной доставке документов, об этом будет сказано в третьем докладе,

Наша совместная работа продолжается.

Проект остается открытым.