

DOI: 10.33581/2020-20-56-64

DSPACE – ОПЫТ САМОСТОЯТЕЛЬНОГО АУДИТА И ОБНОВЛЕНИЯ СЕРВИСА С НЕБОЛЬШОЙ ВНЕШНЕЙ ПОДДЕРЖКОЙ

DSPACE – EXPERIENCE OF SELF-AUDITING AND SERVICE UPDATES WITH LITTLE EXTERNAL SUPPORT

Ефимов Александр Александрович – заместитель начальника отдела информационно-аналитического сопровождения Уральского федерального университета (Россия), e-mail: alexander.efimov@urfu.ru

Efimov Alexander – deputy of head, Department of information and analytical support, Ural Federal University (Russian Federation), e-mail: alexander.efimov@urfu.ru

Аннотация: В статье описывается опыт обновления Электронной библиотеки Белорусского государственного технологического университета (БГТУ), произведенный собственными силами сотрудников БГТУ с методической поддержкой внешнего эксперта. Процесс обновления разделен на этапы, включающие в себя аудит сервиса, устранение проблем, обнаруженных в процессе аудита, планирование обновления, непосредственно обновление, фиксацию результатов, создание отчета.

Abstract: The author describes the experience of updating the institutional repository of BSTU. The update was carried out by the university staff with a little support from an external expert. The upgrade process is divided into stages such as audit, problem solving, upgrade planning, upgrade, report generation.

Ключевые слова: DSpace, обновление DSpace, аудит, качество контента, качество метаданных, тайм-менеджмент, работа в команде.

Keywords: DSpace, upgrading DSpace, audit, content quality, metadata quality, time management, team management.

DSpace – это открытое, свободное кроссплатформенное приложение для создания институционального репозитория. Подразумевая долгосрочное накопление и хранение информации в такой консервативной организации как университет, мы автоматически получаем достаточно консервативный сервис, отражающий научную, публикационную и иные активности университета. При этом, не смотря на понятные требования стабильности, преемственности, защищённости и пр. к сервису также предъявляются требования и по актуальности механизмов агрегирования контента, удобству редактирования данных и метаданных, эффективному представлению данных и метаданных. Всё это вызывает необходимость время от времени пересматривать методики работы с сервисом, проводить обновление как сервисного программного обеспечения, так и операционной системы, проводить аудит контента.

Несмотря на наличие отличной документации на английском языке и очень дружного интернационального сообщества пользователей, порой трудно реализовывать данные задачи полностью самостоятельно и в рабочем порядке, да еще и с минимальными простоями сервиса или вовсе без них. Данная статья описывает успешный опыт реализации описанных выше задач, который позволил сотрудникам, поддерживающим репозиторий, приобрести необходимые навыки, обнаружить проблемы, нуждающиеся в исправлении, уточнить методику для разных категорий пользователей, лучше понять и документировать сервис.

Описание процесса разделено на три больших части, каждая из которых, в свою очередь, разделена на этапы. Части и этапы определены не формально, что лучше отражает суть.

Часть первая – общение с человеком

1. Анализ и формализация ощущений эксплуатанта о текущем состоянии сервиса

Обычно эксплуатант может указать на реальные проблемы, но не может верно определить их источник. Тут помогает и собственный опыт эксперта и дополнительное общение с эксплуатантом. А когда оказывается, что проблемы в общем понятны и как их решать – тоже понятно, у всех добавляется уверенности. На самом деле, показать ошибки, связанные как с контентом, так и с проблемами безопасности, можно на примере любого репозитория. Но к счастью, показать способ устранения ошибки, обычно также легко и быстро, как и указать на ошибку.

2. Анализ и формализация ожиданий эксплуатанта от сервиса в результате аудита и обновления

Обычно эксплуатанты не знают о новых возможностях абстрактной новой версии ПО, но знакомы с опытом коллег. И если у кого-то что-то работает, и если оно работает хорошо – рано или поздно, такой опыт распространяется на всё сообщество. На этом этапе полезно объяснить эксплуатанту о ключевых новых функциях новых версий, о закрытии уязвимостей и пр. Это может повлиять в дальнейшем на методику использования сервиса и сделать сервис более полезным, а его работу более эффективной.

В целом, коллеги, использующие DSpace, частенько игнорируют регулярное извлечение полных текстов (улучшает поиск), обновление OAI фида (обеспечивает отражение в большинстве отраслевых агрегаторов), обновление sitemap (обеспечивает индексацию в поисковых системах). Также, не редки недостоверные handle prefix и некоторые другие базовые настройки. Причем такое положение вещей обычно наблюдается с момента установки сервиса, т.е. всё время его работы. Важно отметить, что сказанное выше не относится к БГТУ, это скорее обобщённая картина.

Часто эксплуатант предлагает просто настроить сервис по результатам аудита и ничего не обновлять, но опыт работы с БГТУ подарил хороший пример необходимости обновления. Команда `dspace oai import -f` на верно настроенном DSpace 6.0 выполнялась около десяти часов, а на DSpace 6.3 на том же объёме фонда – около десяти минут.

3. Уточнение компетенций сотрудников (администраторов, контент-менеджеров, бюрократов) на стороне эксплуатанта, формальное распределение ролей и зон ответственности

Всегда есть «местный админ», «общий админ», «контент-менеджер», «бюрократ» и некоторые другие сотрудники, отвечающие за свои части работы сервиса. Работа со всеми должна быть прозрачна. Именно эти люди берут на себя значительную часть взаимодействия с внутренними службами (IT, PR, Library, etc.) и извне на это не повлиять, поэтому каждый должен быть компетентен и понимать свои задачи, свою ответственность. Также здесь очень важны доверительные человеческие отношения. Но при взаимодействии в некоммерческом формате именно так и бывает. Иначе просто не получается взаимодействие.

4. Создание эффективных каналов связи

Не трудно подобрать такой мессенджер, который будет удобен для общения как в группе, так и личном, передаче файлов, скриншотов, документов и пр. Проблема может возникнуть только в желании или нежелании общаться. Но, как показывает опыт, если дело дошло до обновления, то все достаточно мотивированы и могут и будут говорить, обсуждать и пр.

Часть вторая – общение с компьютером

1. Создание тестовой зоны (одной или нескольких) для проверок и тестов без простоя сервиса

В случае с БГТУ процесс был основан на доверии. Копия (обезличенная без персональных данных) виртуальной машины была

развернута как на мощностях «внешнего консультанта», так и на мощностях ВУЗа. Все подготовительные и тестовые операции проводились на тестовых зонах, обсуждались, делались выводы. Виртуализация в целом, и возможность создать достаточное количество копий, и раздать их всем заинтересованным (в частности), очень сильно развязывает руки и повышает удобство.

2. По результатам аудита: создание списка проблем, которые могут (должны) быть устранены без значительного простоя и до обновления сервиса

Обычно в процессе аудита продуктивного сервиса и различных операций на тестовой зоне появляется список задач, которые необходимо реализовать быстро, т.к. временные штрафы при их устранении велики. Например, исправить недостоверный robots.txt, исправить или просто создать достоверный sitemap, проверить и (при необходимости) исправить какие-то настройки сервиса, влияющие на безопасность (доступный извне solr, например), или адекватную работу сервиса (недостоверный handle prefix). Даже если дальнейшие работы по обновлению не будут проведены, либо будут запланированы на более позднее время, есть такие вещи, которые нужно исправлять здесь и сейчас, т.е. до или независимо от обновления.

3. По результатам аудита: создание списка проблем, которые будут решаться после обновления сервиса в рабочем порядке

Данный момент субъективен, но как показывает практика, чем актуальнее версия DSpace, тем лучше работает извлечение полных текстов, тем быстрее работает обновление индексов solr и пр. Конечно, провести многие (по сути регламентные) процедуры можно и до обновления, но после обновления их снова придётся проводить (с ключом -f), затирая прошлые результаты. Что касается правки (пакетной) метаданных на частые ошибки (такие, как двойные пробелы, пробелы перед запятыми, латинские символы в инициалах авторов с «русскими» фамилиями и пр.) – этот процесс

можно делать всегда... и ошибки будут всегда. Старые исправятся, новые появятся. Процесс этот бесконечный, но и в нём нужна системность.

4. Построение плана обновления сервиса, тестовое обновление на тестовой зоне, анализ

Сама по себе процедура обновления DSpace достаточно проста. Она описана на WIKI ресурсе производителя, в сети также есть альтернативные текстовые и видео-инструкции по обновлению. Сам по себе процесс не сложен, но, это если устанавливался сервис по каким-то правилам, человек, устанавливавший сервис, доступен, и нет никаких проблем или неизвестных.

В БГТУ обновлению сервиса предшествовало обновление операционной системы. Были рассмотрены: вариант с переносом сервиса в виртуальную машину с новой версией ОС, и последующее обновление; и обновление ОС с установленным сервисом. В какой-то момент времени у «внешнего консультанта» было пять экземпляров сервиса в пяти виртуальных машинах. Учитывая положение вещей и ресурсы на месте, было принято решение обновлять ОС с сервисом, «чистить» ОС после обновления, настраивать ротацию логов и пр. и только потом обновлять сервис. Вопрос о том, оптимально ли это решение, не стоит. Факт в том, что решение реализовано, операционная система, установленная в виртуальной машине, поддерживается и получает обновления, сервис обновлён и работает так, как ожидается. Да, «внешний консультант» что-то бы сделал иначе, но специалисты БГТУ, исходя из своих возможностей выбрали другой вариант и реализовали его.

Относительно мягкие установленные сроки позволяли документировать каждый шаг, записывать действия на видео, пробовать, повторять, обсуждать. Напомню, речь идет об обновлении тестовой зоны.

5. Формализация опыта обновления тестовой зоны, уточнение плана обновления, выводы

Опыт обновления тестовой зоны максимально приближен к процессу обновления боевого сервиса. В случае с БГТУ разница была лишь в точке публикации. Процесс был описан, выводы сделаны, опыт формализовался в сценарий для обновления боевого сервиса. Сотрудники БГТУ, участвовавшие в реализации задачи, сами понимали все этапы процесса обновления, понимали кто и что будет делать перед обновлением боевого сервиса, во время обновления, и после. «Внешний консультант» же был на подхвате.

Часть третья – реализация обновления

1. Создание резервной копии

В данном случае проблема резервного копирования решилась посредством создания копии виртуальной машины. Как перед обновлением прикладного программного обеспечения и операционной системы, так и перед обновлением сервиса.

2. Собственно, обновление

Всё прошло так, как задумывалось. А именно, сервис и ОС были не только обновлены с минимальным временем простоя, но были обновлены своими силами без привлечения помощи извне. Команда, поддерживающая репозиторий, понимала каждый этап реализации процесса обновления и каждый этап реализовала. В принципе, всё прошло так, как и планировалось, а «Внешний консультант» был привлечен лишь для того, чтобы констатировать отсутствие проблем после обновления.

Вообще, это самое замечательное, когда в процессе реализации какой-то задачи, не только реализуется сама задача, но и команда, задачу реализовавшая, получает опыт, достаточный для последующего тиражирования технологии.

3. Проверка работы агрегаторов, интеграции с системами мониторинга и поисковыми системами

DSpace – платформа синдикатная. Обычно регистрация во всяческих агрегаторах либо была сделана на старте работы сервиса, либо не была сделана вовсе, а обновленный сервис (без ошибок в OAI, sitemap, robots.txt, с включенным https и пр.) – это хороший повод проверить интеграцию с агрегаторами, поисковыми системами и пр. Опыт БГТУ показал, что проблемы конечно есть, но также было показано, что неразрешимых проблем нет. Теперь всё работает так, как ожидается. Каждый шаг описан, каждый шаг понятен.

4. Проверка нового функционала

Здесь имеется ввиду не столько новый функционал ПО, сколько функционал, который ранее не использовался эксплуатантом: автоматический запуск некоторых регламентных процедур, использование инструментов, которые ранее не использовались и пр. Это может быть как банальная автоматическая регулярная генерация sitemap, так и достаточно тонкая настройка потребления памяти сервисами solr или СУБД. Этот момент, в общем, всё еще пусконаладочный, и его не всегда получается полностью проверить на тестовой зоне, т.к. она банально может быть запущена на ином гипервизоре с иным пулом ресурсов.

5. Формализация опыта обновления боевой зоны, уточнение плана обновления, выводы

В случае с БГТУ, насколько мне известно, всё прошло штатно и серьезных изменений в выводах к обновлению тестовой зоны не было. Была проведена дополнительная работа по настройке точки публикации, но это уже специфика реализации точек публикации веб-ресурсов в конкретной организации. А ваш репозиторий опубликован по https? А какой http-proxy вы используете? А в логах вашего Tomcat реальный адрес клиента? А каков максимальный размер post/get запроса у вашего сервиса?

Все вопросы с подвохом, а в БГТУ всё так, как надо, но, «внешний консультант» реализовал бы иначе.

6. Возврат к списку проблем, устранение которых было решено провести в рабочем порядке после обновления сервиса

Насколько мне известно, коллеги и до обновления репозитория понимали, какой объём работ нужно выполнить для того, чтобы контент репозитория был более репрезентативен, а в процессе это понимание только улучшилось. Работа по редактированию метаданных ведется, внесены изменения в методику работы контент-менеджеров, которые позволяют исключить риски появления некоторых типов ошибок, выполняются проверки. Система работает.

Заключение

В заключении мне хотелось бы поблагодарить коллег. За мотивацию, за компетентность, за профессионализм, за доверие. И пожелать коллегам успехов.