



JISC Project Plan for Project Fedorazon

Overview of Project

1. Background

One of the primary barriers that have been encountered across the UK HE and FE repository community is the lack of technical knowledge in the departments that are most apt to take on repositories. A specific user case -that is often cited- is the willingness by libraries to take on the fundamental role of digital content management, but do not have the hardware nor expertise of how that hardware can scale and maintain its robustness. In addition, where some libraries might have the technical know-how, they do not necessarily have the facilities to host multiple physical servers nor the budget to do so (The Bloomsbury College Libraries are particularly aware of this problem given the cost of hosting local servers in WC1 space). This fundamental hardware computing requirement must be addressed if the 14 million pounds that JISC has invested is not to be drained by the ever increasing need for more servers as more content is collected. In other words, the complexity of physical hardware set-up and maintenance must be hidden away from the institution so true focus can be placed upon enhancing repository content.

2. Aims and Objectives

The Aim of project Fedorazon is to enhance the content of repositories throughout the UK's HE and FE sector by providing solutions for the scalability of repositories as they grow in size and complexity. As a rapid innovation project, it looks to remove the "hardware" barriers involved in launching and maintaining a repository. It will accomplish this initially by enabling the use of Fedora Commons repository software on-top-of Amazon's virtual servers (EC2 & S3). By pre-configuring these servers, any HE/FE institution can "rent" Amazon server space and launch their own secure Fedora repository without having to pre-configure a local server within their institution. In short, institutions can launch their repository service in the same day they decide to have one, and without hiring a "hardware" expert.

Objectives:

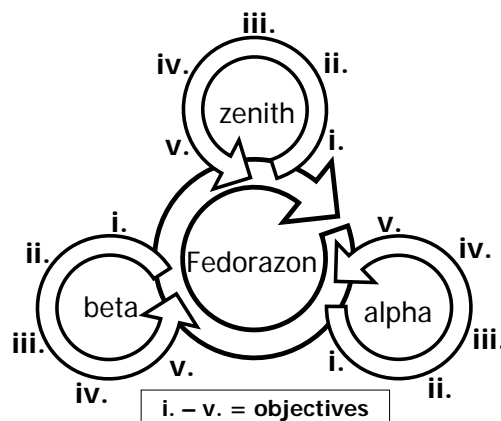
- i.) Provide OS code and "how-to" documentation for using Fedora with EC2, S3 and other services*
- ii.) Encourage and disseminate the use of Fedora on AWS to JISC projects and communities*
- iii.) Provide training on how to deploy Fedora as a component of other services*
- iv.) Support any and all project wishing to deploy repository services on EC2 and S3*
- v.) Provide robust testing and validation of these services as part of the e-Framework & JISC SOA communities.*

3. Overall Approach

The definitive 'aim' of the project is towards the testing and use of specific services (Fedora, EC2 and S3). However, the methodology of this project speaks towards the larger goal of Service Oriented Architecture. We do not know if Amazon will be around in twenty years nor do we know that Fedora will (or that the two of them will be easily compatible) however we do believe that new technologies and services will continue to appear and that Amazon services or Fedora services will be able to be combined with new services to create new applications and useful processes. Therefore, our methodology is intended to demonstrate that services (regardless of proprietary or open source state) should be easily and effortlessly put in place as well as 'swapped-out' using open standards. By taking advantage of SOA in both the education and business sectors, JISC has greater odds of

success as we create our own well defined services. Therefore the project will not only support the Fedora community in establishing a loosely-coupled approach to their own architecture, but also encourage other repositories communities to incorporate AWS services (and other services) to better enable their architecture as it grows and changes.

The means of transforming AWS and Fedora as JISC SOA services are laid out below as a *rapid innovation project*. The method for assuring rapid innovation is achieved will be through a 'rapid prototyping model' of "right – rapid – rough" as used by IDEO Design Company¹. The focus of the project will be upon the delivery of a real world service that will become available in three iterative releases: 1.) alpha (installation of Fedora on EC2), 2.) beta (installation of Fedora on EC2 and S3), and 3.) zenith (swapping out of service components with other services, e.g. Fedora on Local-server-service and S3). During each of these three releases the objectives of the project will be cycled through to assure our overall aim is achieved:



Project Plan Release Cycle ("right-rapid-rough")

For each cycle of the project we will document the use of the services we are coupling according to the JISC e-Framework. This will be done to assess the interoperability of each service within the overall e-infrastructure. Therefore, as we work with each service we will be able to define it according to a common methodology for SOA that will result in greater use and ease by the whole of the JISC community.

The success of the project will be reliant upon the ease of combining these services so that the work of "hardware and systems administration" is abstracted away from the expertise required to administer repository software. In other words, a successful use case for this project would be: and average tech-savy librarian is able to follow the instructions for setting up Fedora on AWS without having to employ a "systems administrator" to do the work of set-up. This includes not having to worry about: scaling, robustness, cost effectiveness, legal constraints or service level agreements. And of course the librarian ("repository administrator") must also trust this service with the institutions content!

4. Project Outputs

Knowledge and Experience Deliverables:

- R&D reports on repository scaling
- R&D reports on repository robustness
- R&D reports on cost effectiveness in "hardware" for repositories
- R&D reports on legal implications for using "virtual servers" (aka utility computing)
- R&D reports on Service Level Agreements for Utility Computing
- Training for Repository Administrators on findings

Technical Deliverables:

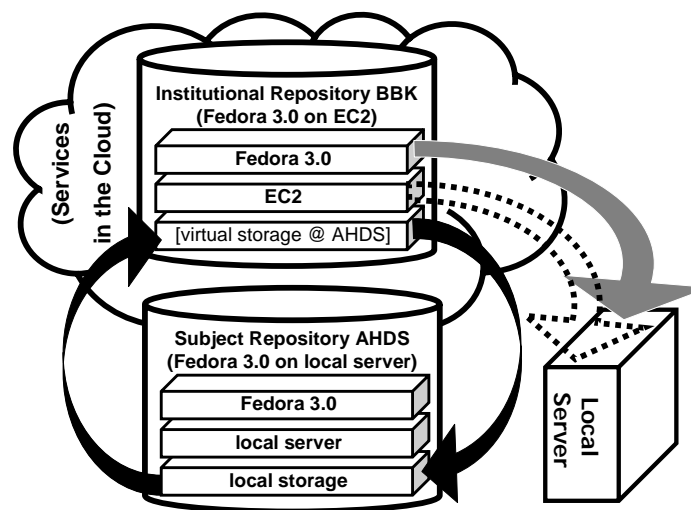
- E-Framework Documentation for each of the services we use, i.e. Fedora, EC2, S3...?
- Open source code for any processes we create to couple services

¹ <http://theartofinnovation.com/>

- User friendly “How-to” Guides on setting up Fedora on EC2 and S3
- Face-to-face training on how to configure above services.

5. Project Outcomes

The best possible outcome of the project would be the ability by JISC institutions to forego the need to have technical administration for hardware servers. This would leave repository staff to focus on the real work of the repository: content population and enhancement. However, the overall outcome of this project needs to be the education of repository managers in regards to their options for hardware for their repository (knowledge and experience deliverables above). Once repositories do become the de facto place to put institutional content the requirements by the repository to scale robustly is paramount. Imagine the scenario where an HE institution creates a tool that will allow all essays and dissertation to be uploaded to the institutional repository for course submission. If the “servers” crash upon promoting this workflow, it could be a devastating blow to the repositories reputation as a “reliable technology”. This project –above all- will aim to demonstrate the importance of being ready for this future repository landscape. This includes being aware of “hardware” services such as EC2 and S3, but also aware that these services are “swappable” with other services (technical deliverables above). The e-Framework will provide the kind of SOA documentation necessary to assure that the technologies we are working with in this project are contributing to the overall JISC community. For the technology literate wing of the JISC community EC2 and S3 (and the rest of AWS services) might well play a part in their repository stack, however it is most likely that these services will be used in unique and different ways. Therefore, the agnostic documentation of all the services used in this project will be the greatest long-term benefit as SOA technicians envision new and interesting ways to use S3, EC2 and Fedora!



“Swappable” service components

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
Repository Managers at JISC institutions	Lower cost in managing repository “hardware”	high
JISC SOA Technologists	Defining of services used in accordance with e-Framework	high
Bloomsbury Consortium Librarians	Use of EC2 and S3 as replacement for physical hardware servers	High
Fedora UK&I User Group	Plausible uses of AWS services with Fedora	Med
Other OS Repository Communities	Adaptation of EC2 and S3 to be used with other repository services (e.g.ePrints, DSpace)	Low

Vendors of Repository Services and Support	Competition to out-sourced server farms or technology support companies (systems admin consultants)	Low
--	---	-----

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing Loss	1	1	1	All staff are replaceable (lots of people use AWS)
Services do not couple as cleanly as hoped	2	3	6	Define the services explicitly so they can be coupled with other services
Service Level Agreements	5	1	5	Watching brief on SLA for Amazon and other Utility computing companies.
Cost of running services	2	3	6	Benchmark use of services and forecast different usage scenario for overall usage
Legal	3	2	6	The legalities of running a service using US legal legislation requires the listing of risk management (watching brief and reports)

8. Standards

Name of standard or specification*	Reference (*for version use Internet Archive and date of this report)
HTTP	http://www.w3.org/Protocols/
ReST API (Fedora)	http://www.fedora.info/wiki/index.php/RESTful_Fedora_Proposal
ReST API (S3)	http://docs.amazonwebservices.com/AmazonS3/2006-03-01/RESTAPI.html
EC2 AMI	http://www.aws.amazon.com/ec2
MySQL 5.x	http://www.mysql.com/
WebDAV / DavFS	http://www.webdav.org/
SSH / SSL	http://www.openssh.com/
Pound HTTP Reverse Proxy	http://www.apsis.ch/pound/
Apache Tomcat	http://tomcat.apache.org/tomcat-5.5-doc/index.html

9. Technical Development

As stated above, this project takes an agile methodology in regards to technology development. Specifically the use of ReST based protocols and best practice in SOA enables the project to change as it moves through its three development cycles. In this way the project can quickly adapt to take on new technologies, all the while documenting the steps it has taken. It is the opinion of the project team that mistakes are as important to document as successes (and perhaps more important from an overall JISC financial point of view). Accordingly, it will do its best to report the accurate state of technology and not drive the technology based on 'commercial sales hype' (though we will not be held accountable for our general excitement about this stuff!).

10. Intellectual Property Rights

Any and all code produced will be published under the BSD Open Source License. In working with proprietary companies will strictly adhere to using open standards well defined boundaries so services can be interoperably replaced.

Project Resources

11. Project Partners

The Bloomsbury Colleges are an established consortium who have numerous project, initiatives and a research centre ongoing throughout Bloomsbury. As part of the established Bloomsbury Learning Environment this project will help contribute to bringing the VLE and Repository services closer together to improve our own internal SOA.

Our lead technology partner is the Centre for e-Research at King's College London who specialise in the knowledge of grid architecture and utility computing. They will help us in the zenith cycle of the project where we will examine what other services we can couple with Fedora, EC2 and S3.

The configuration of Fedora, EC2 and S3 will be outsourced to MediaShelf (contact Matt Zumalt) who will carry out the work of each cycle based on a set of workpackages and specification as recommended by the project team and Bloomsbury Librarians.

12. Project Management

The project is based at Birkbeck Library where the head of the library (Philip Payne) oversees new developments in library technology. The project manager is based in the library systems team in Bloomsbury where the resources and technical support for the project are ample with six college partners supporting this work. The project manager (David F. Flanders) –with experience in several other JISC and HEFCE projects- will oversee this project with 20% of his time. He will oversee the project and produce the majority of reports and research documentation. He will also oversee the assignment of any technological workpackages that will be assigned to MediaShelf (Matt Zumalt) and King's College (Mark Hedges). Technical documentation will be edited by the project manager, but in the first instance written by the technology team member undertaking the work. Contact details for the above mentioned are:

- Philip Payne (Head Librarian), Birkbeck College Library p.payne@bbk.ac.uk
- David F. Flanders (Project Manager), Bloomsbury Colleges, d.flanders@bbk.ac.uk
- Mark Hedges (Deputy Director), King's College e-Research Centre, mark.hedges@kcl.ac.uk
- Matt Zumalt (Owner), MediaShelf Ltd., matt.zumwalt@yourmediashelf.com

13. Programme Support

- An enhanced method for advertising project disseminations (training events, worthwhile documentation, etc), e.g. an RSS feed for all JISC projects events that are taking place?
- More community workshops on e-Framework documentation (not novice how-to but working groups who can sit down together and see how each of us are getting on with it on our own.

14. Budget (Appendix A)

The primary expenditure of the budget is in hardware for testing of utility computing services, specifically Amazon Web Services. The secondary largest expenditure is on staff which will accomplish the programmatic work and documentation necessary to enable AWS-Fedora service as an out of the box configured environment. The institution is contributing primarily to the hardware budget as these will be the most benefit to the institution beyond the life of the project. Detailed budget available in Appendix A.

Detailed Project Planning

15. Workpackages (Appendix B)

For detailed workplan please see appendix B. The following tables list milestones and deliverables:

Alpha Cycle (Oct-Nov 2007):

<u>Technical Workpackages:</u> -Installation of Fedora on EC2 -Set up default Image on EC2 -Test Environment -Deploy Fedora - defaults -Configure Fedora for tBC consortium ² -Scope custom AMI(s) for UK&I Fedora community	<u>Documentation and Training Workpackages:</u> -Help Guide Documentation -eFramework Documentation -Scoping report on roundtable discussion -Download Authentication (Security) -Web Site and Skype for Remote Training -Scoping report on e-Framework participation
---	---

Beta Cycle (Dec-Jan 2007/8):

<u>Technical Workpackages:</u> -Installation of Fedora on EC2 and S3 -Set up default Image on EC2 -Configure Environment -Configure Fedora for tBC consortium -Scope custom AMI(s) for Fedora community	<u>Documentation and Training Workpackages:</u> -Budget Report on Cost of Running Fedora on AWS -Security Report on authorization procedures of AWS -E-Framework Service Expressions Report -Evaluation report of scoping installation on other repository (EPrints/DSpace)
--	---

Zenith Cycle (Feb-March 2008):

<u>Technical Workpackages:</u> -Test EC2 as a switchboard to local storage -Testing AHDS switchboard service -Collect test objects/content -Populating test instance -Porting to Local Server -Testing of Port to Local Server	<u>Documentation and Training Workpackages:</u> -Developer/System Administration Documentation -Deployment Guide & Configuration Report -Help Guides and Installation Instructions (AHDS and Porting) -Updating Public Fedora Docs and wiki -Face-to-Face Training: AWS/Fedora (OR2008 Conference) -Final JISC project report and sign-off
--	--

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Nov 2007	Configuration Environment for Fedora	Which database software and operating system is most likely used for Fedora?	Email Fedora listserv and ask what environment is currently most used	Verbose response from listserv .
Dec 2007	Testing of security and robustness of system	How much content can the system handle and how robust is the service of that content?	Test by programmatically moving content in and out of system	System does not fail, or failure is resolvable. 99% up time.
Jan 2008	Cost analysis of using Fedora on AWS	How much does it cost to use AWS given different needs of JISC repositories?	Cost analysis chart and spreadsheet to enable forecasting of cost.	Cheaper than running local server and hardware for repository.
Feb 2008	Evaluating other service couplings with AWS	Which are some other ways of configuring EC2 and S3 with other services to support JISC UK HE institutions?	Email SUE listserv for response on what other ways their system architecture is configured.	Verbose response by JISC projects and volunteer partner willing to take a new architecture forward.

² OAI-PMH, XACML, LDAP, PostgreSQL, Mulgara, SSL-apache

March 2008	Testing alternate hardware configurations	Is the new set of services robust? How do they compare to the previous configuration?	Programmatic testing of system and comparison to original config (see above)	System does not fail, and is comparable to original config (see above)
April 2008	Training and Documentation Impact	Has anyone else taken up utility computing and was it from advice by project Fedorazon?	Listserv throughout UK Repository Community	Ongoing evaluation and feedback beyond scope of funding.

17. Quality Plan

Output					
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Dec 08	How-to documentation for setting up Fedora & AWS	Testing with user community	Support and use by community for documentation	Project Manager	TBC
Jan 08	Robustness and usability of AWS w/ Fedora	Programmatic testing of content push and pull	System does not fail	MediaShelf	TBC
Mar 08	Usability of S3 and EC2 with other services	Submission to e-Framework for testing and robustness	Monitoring of use of S3 and EC2 via e-Framework website	King's College	TBC

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Jan 08	Face-to-Face training on use of AWS with Fedora	JISC UK HE	Awareness	What these services have been used for and what they could be used for.
April 08	Article and Presentation for OR2008	HE/FE	Awareness	What we've done, what can be done
Ongoing	Website	ALL	Knowledge	This is what has gone on in this project.
Dec 08	e-Framework Documentation	HE / FE	Technical Know-How	This is how these services can be used with other services

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Amazon Machine Image	Registered with AWS AMI registry	Make public in registry
How-to Guides	Website and training outputs	Publish in Jorum
Reports	Listservs and word-of-mouth	Preserve in Repository

Appendixes

Appendix A. Project Budget

Total Expenditures:

Directly Incurred Staff	Year Oct 07- March 08	TOTAL £	
Total Directly Incurred Staff (A)	£11964	£11964	
Non-Staff			
Travel and expenses	£2000	£2000	
Hardware/software	£5000	£5000	
Dissemination	£500	£500	
Evaluation	£500	£500	
Other: Amazon Web Service Rental of EC2 and S3	£4000	£4000	
Total Directly Incurred Non-Staff (B)	£12000	£12000	
Directly Incurred Total (A+B=C) (C)	£23964	£23964	
Directly Allocated			
Estates	£465	£465	
Directly Allocated Total (D)	£465	£465	
Indirect Costs (E)	£2918	£2918	
Total Project Cost (C+D+E)	£27347	£27347	
Amount Requested from JISC	£20000	£20000	
Institutional Contributions³	£7347	£7347	
Percentage Contributions over the life of the project	Institution: X %27	JISC X %73	Total 100%

³ If the institutional contributions include a contribution towards the direct costs of the project please complete a table along the lines of the example overleaf
Page 8 of 12
Document title: JISC Project Plan
Last updated: April 2007

Nature of Institutional Contributions:

Directly Incurred Staff		
Post, Grade & % FTE	£0	£0
Directly Incurred Non Staff		
Hardware/Software etc.	£3964	£3964
Directly Allocated		
Staff, Estates etc.	£465	£465
Indirect Costs		
Indirect Costs	£2918	£2918
Total Institutional Contributions	£7347	£7347

Project Name: Fedorazon
 Version: 1.0
 Contact: d.flanders@bbk.ac.uk
 Date: 30 October 2007

Appendix B. Workpackages

WORKPACKAGES	Month	1 Oct 07	2 Nov	3 Dec	4 Jan 08	5 Feb	6 Mar
1: Project Management							
2: Alpha Cycle and Release							
3: Beta Cycle and Release							
4: Zenith Cycle and Release							
5: Training and Dissemination							

Project start date: October 2007 / **Project completion date:** April 2008 / **Duration:** 6 months

Workpackage and activity	Earliest Start Date	Latest Completion Date	Outputs (Clearly indicate deliverables and reports in bold) Milestone = *	Respo nsibilit y
WORKPACKAGE 1: Project Management <i>Objective: This WP encompasses all planning and reports to be written along with project documentation and overall guidance of the project from start to finish.</i>			<ul style="list-style-type: none"> Overall management of project and tracking of project progress Start-up report Completion report and sign-off Ongoing updates via wiki and blog Communication with JISC programme via listserv and meetings Watching briefs and assurance of project documentation completion 	
1. Project Planning	Oct 07	Mar 08		
2. Project Report Writing	Oct 07	Mar 08		
3. Communication w/ JISC community	Oct 07	Mar 08		

WORKPACKAGE 2: Alpha Cycle <u>Objective:</u> Installation of Fedora on EC2			<ul style="list-style-type: none"> List of configuration variables for Fedora Help Guide Documentation eFramework Documentation Authentication robustness test (security) report Watching brief on Repository Hardware solutions 	
4. Gathering of configuration environment variables from Fedora community	Oct 07	Oct 07		MS
5. Set-up of Amazon Machine Image for Fedora on EC2 based on config variables	Oct 07	Nov 07		
6. Testing of Fedora on EC2 (put/get content)	Nov 07	Dec 07		
7. Documentation on Fedora set-up on EC2				
WORKPACKAGE 3: Beta Cycle <u>Objective:</u> Installation of Fedora on EC2 and S3			<ul style="list-style-type: none"> Config options for EC2 and S3 (i.e. WebDAV) Help Guide Documentation eFramework Documentation Cost Evaluation Report and Spreadsheet Watching brief on Utility Computing options 	
8. Research possible connection via Fedora (on EC2) and S3	Nov 07	Dec 07		
9. Implement WebDAV between Fedora (EC2) and S3	Nov 07	Dec 07		
10. Test put/get on Fedora (EC2+S3)	Nov 07	Jan 08		
11. Documentation of Fedora on EC2 and S3	Dec 07	Jan 08		
WORKPACKAGE 4: Zenith Cycle <u>Objective:</u> Switching out of services for another service (e.g. other utility computing service?)			<ul style="list-style-type: none"> R&D of options for connecting other services Establish test-case for use of other service Scope other service specification (standards brief) Implementation of technology for other standards 	
12. R&D for other services to couple EC2 and S3	Jan 08	Feb 08		DFF
13. Swap out service for other service	Jan 08	March 08		KC

14. Test and compare other service with original services configuration	Feb 08	March 08		
WORKPACKAGE 5: Training and Dissemination			<ul style="list-style-type: none"> • Publish e-Framework documentation • Publish and Disseminate Help Documentation • Publish AMI for Fedora on EC2 and S3 • Disseminate Project Outcomes and Findings • Training session on use of EC2 and S3 	
<u>Objective:</u>				
15. Announce training and set training date for face-to-face training	Jan 08	March 08		
16. Finalise e-Framework Documentation for each service used	Jan 08	March 08		
17. Finalise Help Documentation and Technical Reports	Jan 08	March 08		
18. Announce final documentation and test-cases to JISC community	Feb 08	March 08		
19. Disseminate findings at conference (OR2008) and in appropriate publication (DLib)	Jan 08	March 08		

DFF = David F. Flanders
MS = MediaShelf
KC = King's College