



**SEVENTH FRAMEWORK PROGRAMME
FP7-ICT-2009-6**

BlogForever
Grant agreement no.: 269963

BlogForever: D5.5 Case Studies Comparative Analysis and Conclusions

Editor:	Vangelis Banos
Revision:	First
Dissemination Level:	Public
Author(s):	Vangelis Banos, Silvia Arango-Docio, Ed Pinsent, Patricia Sleeman, Ioannis Manolopoulos, Ilias Trochidis.
Due date of deliverable:	31/5/2013
Actual submission date:	31/5/2013
Start date of project:	01 March 2011
Duration:	30 months
Lead Beneficiary name:	AUTH

Abstract: This report presents the analysis and conclusions of the BlogForever Case Studies based on internal and external testing results, as well as recorded system logs. The case studies are paramount for the evaluation of the performance and the impact of the BlogForever platform.

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)

The **BlogForever** Consortium consists of:

Aristotle University of Thessaloniki (AUTH)	Greece
European Organization for Nuclear Research (CERN)	Switzerland
University of Glasgow (UG)	UK
The University of Warwick (UW)	UK
University of London (UL)	UK
Technische Universitat Berlin (TUB)	Germany
Cyberwatcher	Norway
SRDC Yazilim Arastrirma ve Gelistrirme ve Danismanlik Ticaret Limited Sirketi (SRDC)	Turkey
Tero Ltd (Tero)	Greece
Mokono GMBH	Germany
Phaistos SA (Phaistos)	Greece
Altec Software Development S.A. (Altec)	Greece

History

<i>Version</i>	<i>Date</i>	<i>Modification reason</i>	<i>Modified by</i>
0.1	20/3/2013	Initial draft	Vangelis Banos (AUTH)
0.2	15/4/2013	Second draft	Vangelis Banos (AUTH)
0.3	24/5/2013	Third draft	Vangelis Banos (AUTH), Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent (UL).
0.5	29/5/2013	Fourth draft	Vangelis Banos (AUTH), Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent (UL)
1	31/5/2013	First version	Vangelis Banos (AUTH), Yannis Manolopoulos (AUTH), Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent (UL), Ilias Trochidis (Tero).
1.1	25/9/2013	Extended version	Vangelis Banos (AUTH)

Table of Contents

TABLE OF CONTENTS	4
LIST OF TABLES	5
LIST OF FIGURES	6
EXECUTIVE SUMMARY	7
1 INTRODUCTION	8
1.1 BACKGROUND	8
1.2 WORK PACKAGE 5 TASKS	9
1.3 OBJECTIVES OF THIS DELIVERABLE	10
1.4 CONNECTIONS WITH PREVIOUS AND PENDING WORK	10
1.5 STRUCTURE OF THE REPORT.....	10
2 METHODOLOGY	12
2.1 GENERAL EVALUATION METHOD.....	12
2.2 RESEARCH QUESTIONS AND THEMES.....	13
2.3 ORGANISATION AND SCHEDULING OF THE EVALUATION.....	16
2.4 BLOGFOREVER CASE STUDIES RESULTS INTENDED AUDIENCE.....	16
3 INTERNAL TESTING	18
3.1 METHOD.....	18
3.2 RESULTS.....	20
4 EXTERNAL TESTING	26
4.1 METHOD.....	26
4.2 RESULTS.....	27
4.2.1 CS1 Answers Summary.....	31
4.2.2 CS2 answers summary.....	34
4.2.3 CS3 answers summary.....	35
4.2.4 CS4 answers summary.....	36
4.2.5 CS5 & CS6 answers summary.....	39
5 SYSTEM LOGS	40
5.1 METHOD.....	40
5.2 CS1 & CS2 RESULTS	41
5.3 CS3 & CS4 RESULTS	47
5.4 CS5 & CS6 RESULTS	54
6 CONCLUSIONS	58
6.1 SCORING RESULTS RECAP	59
6.2 RESEARCH QUESTIONS OUTCOMES.....	60
7 REFERENCES	64

List of Tables

Table 1, Research Questions [1].....	13
Table 2, Metrics.....	14
Table 3, Themes and Research Questions Association	15
Table 4, BlogForever Case Studies Schedule.....	16
Table 5, Internal Testing Template	20
Table 6, Summary of scores for each Theme	23
Table 7, External Testing Questionnaire	26
Table 8, Association between CS1 external User Questionnaire questions and Research Themes ..	27
Table 9, External Testing Scores Summary	28
Table 10, Metrics [15]	40
Table 11, Metrics and Themes association [10].....	41
Table 12, Overall usage data for CS1 & CS2 [10]	41
Table 13, CS1-CS2 Metric 5: Number of python code errors.....	45
Table 14, CS1-CS2 Metric 5: Errors grouped by python module	46
Table 15, CS1-CS2 Metric 6: HTTP status distribution.....	46
Table 16, CS1-CS2 Metric 7 Page loading time distribution	47
Table 17, CS1-CS2 Metric 8: Pages per visit.....	47
Table 18, CS1-CS2 Metric 9: Average Visit Duration.....	47
Table 19, Overall usage data for CS3 & CS4.....	48
Table 20, CS3-CS4 Metric 5: Python code errors	53
Table 21, CS3-CS4 Metric 5: Errors grouped by python module	53
Table 22, CS3-CS4 Metric 6: HTTP Status distribution	53
Table 23, CS3-CS4 Metric 7: Page loading time distribution	54
Table 24, CS3-CS4 Metric 8: Pages per visit.....	54
Table 25, CS3-CS4 Metric 9: Average Visit Duration.....	54
Table 26, CS5-CS6 Metric 6: HTTP status distribution.....	55
Table 27, CS5-CS6 Metric 5: Python top code errors	56
Table 28, CS5-CS6 Metric 5: Errors grouped by python module	56
Table 29, CS5-CS6 Metric 7: Page loading time distrubution	57
Table 30, External Testing Scores Summary	59
Table 31, Internal Testing Scores Summary	59
Table 28, RQ3 related Theme scores	60
Table 33, RQ4 related Theme scores	61
Table 34, RQ5 related Theme scores	62

List of Figures

Figure 1, BlogForever Platform Overview	8
Figure 2, Work Package 5 timeline and interactions [10]	9
Figure 3, Visual representation of the scores for each case study	23
Figure 4, Scores grouped by Theme and Case Study	28
Figure 5, Average scores for each Theme, plotted in range of 4 to 5	29
Figure 6, Distribution of scores for T1: Using blog records and T2: System integrity	29
Figure 7, Distribution of scores for T3: Sharing and interaction and T4: Searching	30
Figure 8, Distribution of scores for T5: Access and T6: Data integrity	30
Figure 9, Distribution of scores for T7: Preservation and T8: Functionality	31
Figure 10, Distribution of scores for T9: System navigation and T10: System terminology	31
Figure 11, CS1-CS2 Metric 1: Content records page views.....	42
Figure 12, CS1-CS2 Metric 2: Export page views	43
Figure 13, CS1-CS2 Metric 3: Search page views.	44
Figure 14, CS1-CS2 Metric 4: Achieved goals in GA.	45
Figure 15, General information on visits for CS3 & CS4	48
Figure 16, CS3-CS4 Metric 1: Content records page views.....	49
Figure 17, CS3-CS4 Metric 2, Export page views	50
Figure 18, CS3-CS4 Metric 3: Search page views	51
Figure 19, CS3-CS4 Metric 4: Achieved goals in GA	52
Figure 20, General information on visits for CS5 & CS6	55

Executive Summary

This report presents the analysis and conclusions of the BlogForever platform evaluation conducted through the design, implementation and evaluation of the BlogForever Case Studies. The goal is to present the conclusions of all evaluation methods, namely Internal Testing, External Testing and System Logs, and combine the outcomes, experiences and feedback into a meaningful whole.

In order to achieve this, the following evaluation methods are used:

- **Internal Testing** using internal observations and reporting of the implemented features.
- **External Testing** involving third party users who conducted predefined tests and answered questionnaires.
- **System Logs** using web analytics and server logs capturing and processing, while internal and external testing operated.

The outcomes of the evaluation are analysed using quantitative and qualitative methods, aiming to answer the following Research Questions:

- **RQ1:** What are the particular **problems** the implementation is facing? Or are the BlogForever software implementation processes an **overall success**?
- **RQ2:** Are complex BlogForever platform search strategies **working efficiently** when high levels of content are available within the BlogForever platform?
- **RQ3:** How **useful** is the BlogForever platform as a whole?
- **RQ4:** Does the use of the BlogForever repository lead to **successful** results for the different users?
- **RQ5:** How **user friendly** are the BlogForever platform functions for the different designated blog communities?

A set of Themes were also devised as part of the report D5.2 Implementation of the Case Studies [1] in order to further elaborate the Research Questions and try to connect them with the evaluation methods:

- T1:** Using blog records,
- T2:** System integrity,
- T3:** Sharing and interaction,
- T4:** Searching,
- T5:** Access,
- T6:** Data integrity,
- T7:** Preservation,
- T8:** Functionality,
- T9:** System navigation,
- T10:** System terminology.

The outcomes of this evaluation will show that the BlogForever Case Studies have been conducted in a meaningful and comprehensive way. The platform has been evaluated from many different aspects and the results validate and complement the existing development work. In summary:

- **Internal Testing** evaluated all implemented features and resulted in the conclusions that most areas are working moderately well to very well.
- **External Testing** evaluated the BlogForever platform from an outside user point of view, and resulted in positive feedback.
- **System Logs** calculated a set of metrics from the BlogForever servers during the case studies and demonstrated that system integrity and performance is high.

Finally, the conclusions of this report present an evaluation of the performance and the impact of the BlogForever platform which can act as a guide for all project stakeholders.

1 Introduction

The BlogForever project aims to create robust aggregation, preservation, management and dissemination facilities for blogs. An integral part of the development process is the implementation of specific case studies in order to evaluate and test the created infrastructure on extensive and diverse sets of weblogs. In this report, we present the analysis of the case studies. We summarise their results of evaluations and derive concrete conclusions.

In this section, we present some background information on the BlogForever project as well as more details on Work Package 5 (WP5) case studies and validation. Furthermore, we elaborate on the objectives of these deliverables and their connection with previous and pending work.

1.1 Background

The BlogForever project is creating a software platform capable addressing the problem of weblog preservation. This platform can be used by memory institutions, researchers and universities, as well as communities of bloggers to build simple weblog archives for a variety of purposes. The entire BlogForever system comprises two discrete elements:

- the weblog spider which is responsible for monitoring, harvesting and analysing weblogs, and,
- the weblog repository component which is responsible for weblog data preservation, ensuring weblog proliferation, safeguarding blog integrity, authenticity and long-term accessibility over time, and allowing for better sharing and re-using of contained knowledge.

An overview of the BlogForever platform is presented in Figure 1.

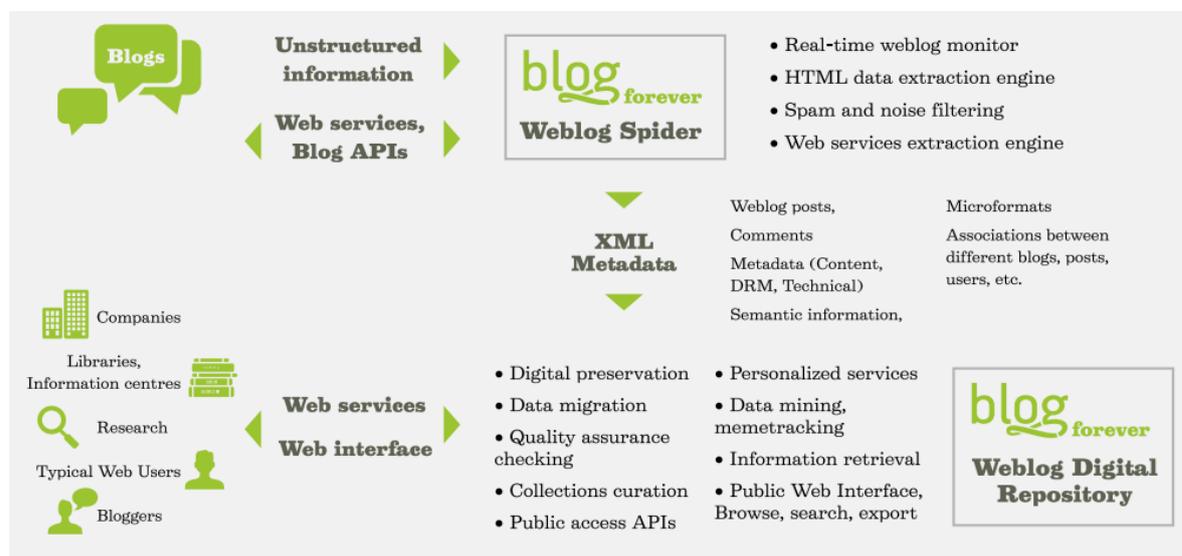


Figure 1, BlogForever Platform Overview

The current status of the BlogForever work grouped by Work Package can be summarised as follows:

- **Work Package 2 (WP2):** the aim was to study weblog structure and semantics. To achieve this goal, WP2 analysed a large number of blogs [2], developed a generic data model for blog structure and archiving and studied interblog relationships [3]. Additionally, WP2 studied weblog ontologies [4] and weblog data extraction [5].

- **Work Package 3 (WP3):** the aim is to study weblog preservation policies. To achieve this goal, WP3 studied blog preservation strategies [6], interoperability prospects [7] and is currently studying DRM strategies.
- **Work Package 4 (WP4):** the aim is to develop the BlogForever software platform. Towards this goal, WP4 compiled the user requirements and platform specifications [8], designed the weblog spider and repository [8] [9] and is currently in the process of implementation.

In the following subsection, we present WP5 in more detail.

1.2 Work Package 5 Tasks

Work Package 5 aims to design and implement specific case studies in order to evaluate and test the created infrastructure on extensive and diverse sets of weblogs. The case studies will be both generic (collecting weblogs from a wide array of topics) and domain specific (for example a case study in University bloggers community). Thus the case studies will provide the required breadth and depth to validate the developed tools, and guarantee that the project's results could be successfully and widely replicated after the project ends. The impact of the digital repository will be evaluated by monitoring system usage and gathering user feedback. Overall, WP5 consists of the following tasks:

1. **Task 5.1 Specification of the case studies:** Design a quantitative software testing approach in conjunction with qualitative criteria to evaluate the BlogForever platform.
2. **Task 5.2 Implementation of the case studies:** Manual and automated collection of data from questionnaires, direct observation, interviews of users and system log files.
3. **Task 5.3 User feedback and evaluation:** Comparative analysis and conclusions.

T5.1 was completed in June 2012. T5.2 and T5.3 are planned to be finished in May 2013. A visual representation of their workflow and interconnection is presented in Figure 2.

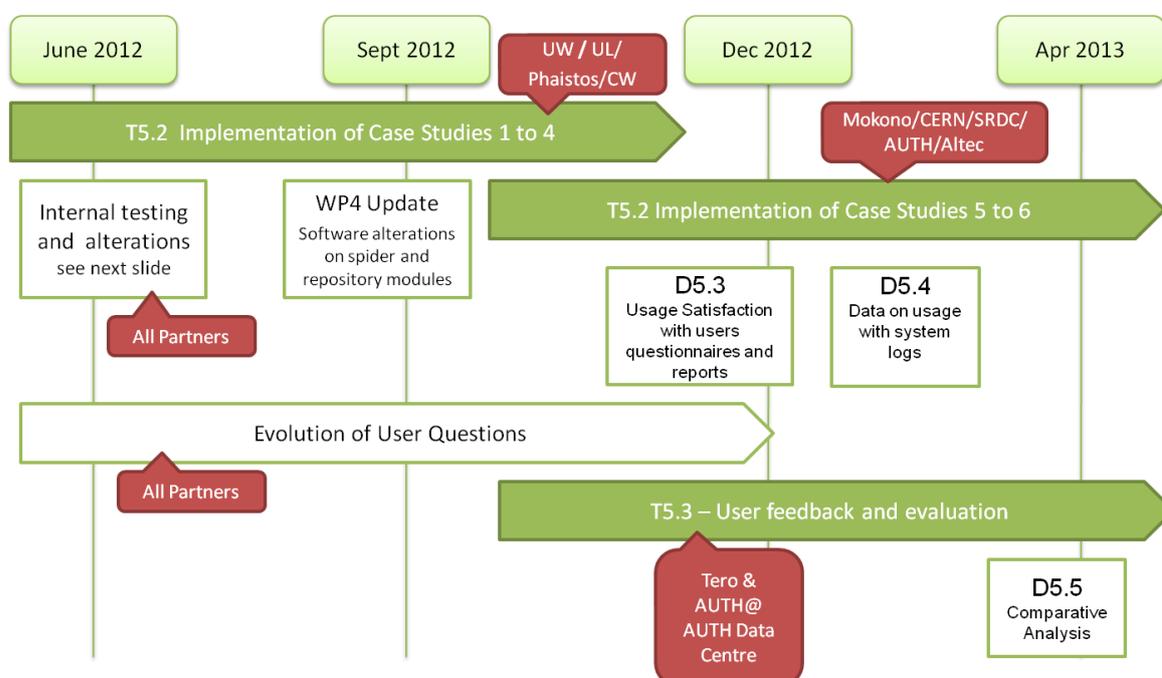


Figure 2, Work Package 5 timeline and interactions [10]

This report is the final step of T5.3 User feedback and evaluation. Its objectives are presented in the following subsection.

1.3 Objectives of this deliverable

This deliverable (D5.5) belongs to the Task 5.3 User feedback and evaluation. The main aspects of Task 5.3 are:

1. Gather user feedback and perform evaluation.
2. Perform analysis on the system logs to closely monitor the application and identify issues.
3. Feedback is given to Task 4.6 (Integration and Standardization), which will be performed in parallel with the current task.
4. Questionnaires, structured interviews and direct observations will be utilized twice during the task, once during its early stages and once at its end. These methods will help the project evaluate the system from the users' point of view, providing valuable feedback, which cannot be captured through system logs.
5. Documentation held for case logs and other project evaluation feedbacks will result into deliverables D5.3 and D5.4.
6. Finally, case study reports are collectively analysed in order to demonstrate the possible courses of action as well as current and expected impact on the blogging community. The results of this will feed into the final D5.5 deliverable.

The objectives of this deliverable are to:

1. Present a summary of the case studies implementation results,
2. Produce guidelines, which will assist the other WPs,
3. Analyse the results and form conclusions.

The connections of D5.5 with previous and pending work are explained in the following subsection.

1.4 Connections with previous and pending work

In short, the case studies comparative analysis and conclusions is based on the following deliverables:

- **Weblog spider implementation deliverables** (D4.3 Initial Weblog Spider Component [9] & D4.6 Final Weblog Spider Component [10]) present the development activities and the features of the weblog spider. Additionally, these reports explain the usage and management of the spider. This information is essential for the execution of the case studies and their analysis.
- **Weblog repository implementation deliverables** (D4.5 Initial Weblog Digital Repository Prototype [11] & D4.7 Final Weblog Digital Repository [12]) present the development activities and features of the weblog repository. Additionally, these reports explain the usage and management of the repository. This information is essential for the execution of the case studies and their analysis.
- **D5.1 Design of Specific Case Studies** [13] presents the design and specifications of the six case studies under analysis in this deliverable.
- **D5.2 Case Studies Reports** [1] presents the internal and external testing results for the BlogForever Case Studies.
- **D5.3 User Questionnaires and Reports** [14] presents the case studies user questionnaires and reports. This information is under analysis in this deliverable.
- **D5.4 System logs** [15] present the case studies system weblogs gathered throughout the implementation. This information is under analysis in this deliverable.

1.5 Structure of the report

The remaining content of this report is divided into five sections. Section 2 presents the general methodology and rationale used for the analysis of the case studies results, as well as their context

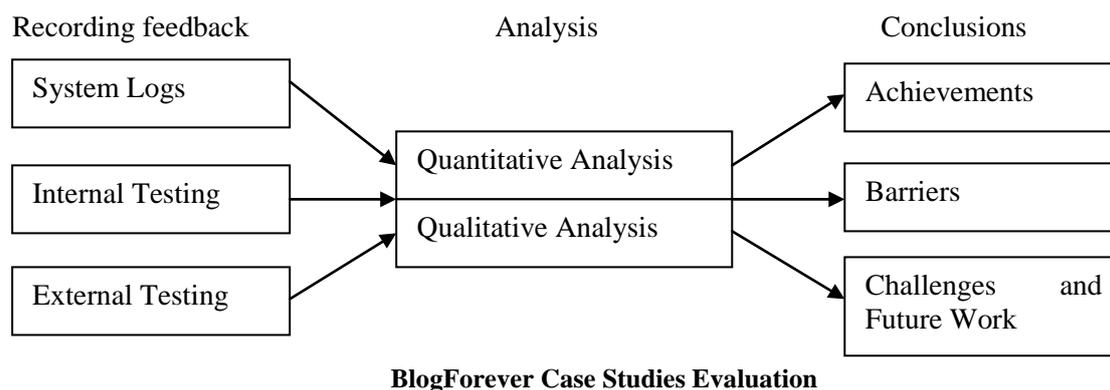
and intended audience. Section 3 presents the internal testing method and results. Section 4 presents the external testing and results. Section 5 presents the System Logs analysis method and results. Section 6 presents the conclusions of the evaluation.

2 Methodology

In this section, we present the methods we used to analyse the BlogForever Case Studies. First, we present the general approach and the key concepts of Research Questions and Themes. Following, we present the evaluation schedule and the intended audience of this work.

2.1 General Evaluation Method

The general method we use to evaluate the BlogForever Case Studies consists of three key steps: a) record feedback from multiple processes, b) analyse it independently, and, c) draw conclusions. An outline of the evaluation method is presented in the following figure:



In order to achieve the best results, three processes were used simultaneously to gather input, in order to provide a better understanding of the purpose, current status, functionality and issues of the BlogForever platform. These are:

1. **System Logs** recorded by the servers running the case studies. Detailed information on the system logs is part of deliverable D5.4 System Logs.
2. **Internal Testing** used to gather feedback from project partners while testing features and recording their status. Specific reports were created for each case study, presenting the outcomes of internal testing. Detailed information on internal testing is part of D5.2 Implementation of Case Studies.
3. **External Testing** used to gather feedback from 3rd party users involved in testing. External users submitted specially designed User Questionnaires. Detailed information on the User Questionnaires is part of D5.3 User Questionnaires and reports.

To analyse recorded feedback, we employ both **quantitative methods** to express in numbers the achievements of project goals as well as **qualitative methods** to interpret user feedback and draw detailed conclusions.

The foundation of the evaluation is a set of **Research Questions** initially defined in D5.1 Design of Specific Case Studies. Moreover, **Themes** have been also devised to assist in rationalising the outputs of user questionnaires, internal tests and system logs. Furthermore, **Metrics** were defined to express software variables defined in system logs. These concepts are elaborated in Section 2.2

The outcomes of the evaluation are not restricted to the scoring achieved for the aforementioned variables. In addition, conclusions will evaluate the performance and the impact of the BlogForever platform. The evaluation conclusions are presented in Section 6 Conclusions.

2.2 Research Questions and Themes

The research design of the BlogForever Case Studies established an initial set of general Research Questions, presented in Table 1. Their role is to guide the WP5 research team members and keep them focused on the evidence and data needed to validate the BlogForever features impact [10].

ID	Research Question
RQ1	What are the particular problems the implementation is facing? Or are the BlogForever software implementation processes an overall success ?
RQ2	Are complex BlogForever platform search strategies working efficiently when high levels of content are available within the BlogForever platform?
RQ3	How useful is the BlogForever platform as a whole?
RQ4	Does the use of the BlogForever repository lead to successful results for the different users?
RQ5	How user friendly are the BlogForever platform functions for the different designated blog communities?

Table 1, Research Questions [1]

The evaluation of the first two Research Questions regarding the implementation and the efficiency of the BlogForever platform are answered by the results of the System Logs evaluation. Using the data gathered from System Logs, we specify some explicit Metrics which are relevant to our Research Questions and Themes. These Metrics are outlined in Table 2.

ID	Metric	Description
M1	Content records page views	Repository pages presenting records (blogs, posts, pages or comments). Example: http://bf3.csd.auth.gr/record/5394?ln=en
M2	Export page views	Repository pages used to export content. Example of record export: http://bf3.csd.auth.gr/record/5394/export/hx?ln=en Example of search results export: http://bf3.csd.auth.gr/search?ln=en&p=document&f=&action_search=Search&c=BlogForever+Test+Repository&sf=&so=d&rm=&rg=10&sc=1&of=xm
M3	Search page views	Repository pages used for search. Example: http://bf3.csd.auth.gr/search?ln=en&sc=1&p=analysis&f=&action_search=Search&c=Blogs&c=Posts&c=Comments&c=Pages
M4	Achieve goals in Google Analytics	Goals are a versatile way to measure how well a website fulfils specific objectives, which can be a set of consecutive actions in the website.
M5	Number of python code errors	The number and nature of python errors is important for the system integrity.
M6	HTTP status distribution	The distribution of HTTP responses provides an insight on application stability and integrity.
M7	Page loading time distribution	Average web page loading time is a characteristic of website performance.
M8	Pages per visit	The more pages per user visit is directly relevant to

		the quality of system navigation and functionality.
M9	Average visit duration	The length of user visit duration is directly relevant to the quality of system navigation and functionality.

Table 2, Metrics

In order to elaborate the remaining Research Questions, a set of ten Themes were defined to help rationalising the outputs of all evaluation and connect them with the Research Questions. The terms used do not relate to any technical or development terms previously used within the project (e.g. when building the platform), and are intended to be as clear and simple as possible, in order to promote the point of view of a user. The Themes and their connections with the Research Questions are presented in Table 1.

Theme	Related Research Questions	Expected results of RQ	Rationale linking Theme to RQ
1: Using blog records	RQ5: How user friendly are the BlogForever platform functions for the different designated blog communities?	Expected results in RQ5: User satisfaction , Usability, Strengths, Weaknesses	The theme “Using blog records” refers to the user experience, satisfaction and usefulness of the archived blogs collections tested within the different versions of the BlogForever repository.
2: System integrity	RQ3: How useful is the BlogForever platform as a whole?	Expected results in RQ3: Platform is sustainable and it meets the users' needs	The theme “System Integrity” covers whether the system is logical and secure. The tests are dependent on the software integrity level or risk level.
3: Sharing and interaction	RQ5: How user friendly are the BlogForever platform functions for the different designated blog communities?	Expected results in RQ5: User satisfaction, Usability , Strengths, Weaknesses	The theme “Sharing and interaction” refers to the ability of BF users to share content and metadata with others, including other users of the platform, and any external use via social software.
4: Searching	RQ4: Does the use of the BlogForever repository lead to successful results for the different users?	Expected results in RQ4: Searched content is found fast and in an organised manner.	In the “Searching” theme, the tests focused on how the platform performs searches , and how users can use and interpret the results of searches.

5: Access	RQ4: Does the use of the BlogForever repository lead to successful results for the different users?	Expected results in RQ4: Searched content is found fast and in an organised manner.	The theme “Access” represents how the platform allows access to the blog records, and how it presents dissemination copies of the content.
6: Data integrity	RQ3: How useful is the BlogForever platform as a whole?	Expected results in RQ3: Platform is sustainable and it meets the users' needs	The “Data integrity” theme assesses if the blog datasets are properly captured, well-maintained and consistent
7: Preservation	RQ3: How useful is the BlogForever platform as a whole?	Expected results in RQ3: Platform is sustainable and it meets the users' needs	The theme “Preservation” provides tests to establish if it is possible to preserve blogs.
8: Functionality	RQ3: How useful is the BlogForever platform as a whole?	Expected results in RQ3: Platform is sustainable and it meets the users' needs	The “Functionality” theme assesses tests related to the functions available to users and administrators.
9: System navigation	RQ5: How user friendly are the BlogForever platform functions for the different designated blog communities?	Expected results in RQ5: User satisfaction, Usability , Strengths, Weaknesses	The “System functionality” theme refers to general navigation aspects of the system.
10: System terminology	RQ5: How user friendly are the BlogForever platform functions for the different designated blog communities?	Expected results in RQ5: User satisfaction for platform usability (strengths and weaknesses)	The “System terminology” theme includes instructions, help pages, and other aspects of terminology in the platform.

Table 3, Themes and Research Questions Association

Table 3 presents connections between all the Themes and RQ3, RQ4, RQ5. The first two Research Questions are not connected. This is happening for the following reasons:

- RQ1 (Are the BlogForever software implementation processes an **overall success?**) relates to specific features from D4.1; since the Case Studies were designed to address these

features, RQ1 implicitly applies to all the themes. For this reason RQ1 is not explicitly included in Table 3.

- RQ2 (Are complex BlogForever platform search strategies **working efficiently** when high levels of content are available within the BlogForever platform?) relates to the scalability of the system when high volumes of content are searched. This stress testing feature is the subject of case study 6, so as of May 2013 case study 6 features have not been implemented or tested to yet link RQ2 to a theme For this reason RQ2 is not yet explicitly included in the table; it will be linked to a theme in the next iteration of the current deliverable in Month 30.
- In the Rationale column, the explicit linkage between a Theme and the expected results from a RQ is further explained and demonstrated by highlighting common terminology between the two in bold type.

The evaluation of the BlogForever platform capabilities in relation to Research Questions, Themes and Metrics will result in concrete findings regarding its performance and impact

2.3 Organisation and scheduling of the evaluation

The original BlogForever Case Studies plan was presented in D5.1 Design of Specific Case Studies in June 2012. Nevertheless, the case studies were conducted later than planned, as presented in Table 4, due to changes in the implementation plan.

<i>1st phase of development</i>		
	Planned	Implemented
Case Study 1	July - August 2012	September - December 2012
Case Study 2	July - August 2012	September - December 2012
Case Study 3	September - November 2012	January - April 2013
Case Study 4	December - February 2013	November - March 2013
<i>2nd phase of development</i>		
Case Study 5	September 2013 - February 2013	July - August 2013
Case Study 6	December 2012 - March 2013	July - August 2013

Table 4, BlogForever Case Studies Schedule

All the evaluation procedures (Internal Testing, External Testing, System Logs) were conducted in parallel with the case studies.

2.4 BlogForever Case Studies Results Intended Audience

The following specialist groups are the intended audience of this report:

- **BlogForever project management**, because the findings will be used to assess user opinions about the platform, guide next development steps and find potential new opportunities.
- **BlogForever developers**, because the case studies' findings provide necessary feedback to address software issues and improve the platform.
- **Potential adopters** of the platform, because the outcomes of the case studies provide extensive information on the strengths and weaknesses of the BlogForever platform, and help decide on its adoption.

- **Researchers** interested in weblog archiving, web archiving, digital repository software platforms and applied research.

The evaluation methods and outcomes for all case studies are presented in the following sections.

3 Internal Testing

In the following, we present the Internal Testing method and results.

3.1 Method

The case studies researchers and software testers carry out the **internal** assessment of the implementation process following the following software testing principles [1]:

- ✓ **Principle - Testing shows the presence of defects, not their absence:**
Testing can show that there are defects. Testing cannot prove that the BlogForever platform is defect-free. Even if no failures are found during testing, this is no proof that there are no defects.
- ✓ **Principle - Exhaustive testing is not possible:**
Every test is always just a sample. The test effort is therefore controlled, taking into account risk and priorities.
- ✓ **Principle - Testing activities should start as early as possible:**
Testing activities should start as early as possible in the software lifecycle and should focus on defined goals. This contributes to finding defects early.
- ✓ **Principle - Test is context dependent:**
Testing must be adapted to the risks inherent in the use and environment of the platform tested. Therefore, no two platforms should be tested in the exactly same way.
- ✓ **Principle - The fallacy of assuming that no failures mean a useful system**
Finding failures and repairing defects does not guarantee that the system as a whole meets user expectations and needs. Early involvement of the users in the development process and the use of prototypes are preventive measures intended to avoid problems.

For each case study, they compute and formalise the assessment data for each set of developed features, using the template presented in Table 5. This case study template summarises different outcomes for each feature available through **internal** observations of the BlogForever implemented platform with the corresponding blogs sampled.

BlogForever Platform Case Study Template:

Aim to use the following template to specify details of the implementation of different software features and details of the case study research objectives for the blog group selected.

1. Reasons for selection of the source blog group:

(Description of the source blogs)

2. Objective of the case study:

(Case study topic, what is the overall research objective)

3. Executive Summary of case study:

(Sources: section 4.b and section 6 of this template)

4. Integration and set up:

a. Description of set up and integration of BF platform:

b. Summary of outcome of integration:

5. Research Question ID (i.e RQ1.1): Title of Research Question

Description of the research question	(If necessary define any sub-research questions more specific than the original research questions presented in section 2.3 of this report - WP5 Research Design)
Requirement original ID (From D4.1)	(From D4.1)
Feature ID (From WP4)	(From WP4)
Detail of potential stakeholders (External) Detail of actors (see details in use case tables in 4.1 report) Unit of analysis	Key players identified: bloggers (blog authors); blog providers; libraries, museums, information systems centres and archives; universities and research institutes; hosting companies and blog readers (journalists and businesses companies). Blog database; blog properties; spider elements; repository elements...
Pre-requisites to enable feature/requirement	Installations (BlogForever platform current setup); configurations; monitoring systems...
Expected Outcome	Pre-testing details of what the research team member expects of the feature
Testing Details	Testing description
Was the expected outcome achieved	(Yes/ No)
Report	Describe actual outcome with specific examples
Does outcome match possible stakeholder expectations?	(Yes/ No plus reasons and/or examples)
Recommended corrective measure	(if appropriate)
Score Functionality	Expected outcome versus actual outcome: 1=Did not work as expected 2=Some areas worked as

	expected 3=Most areas worked as expected 4=All work as expected 5=Worked better than expected
Further analysis needed	Log file needed and/or questionnaire statement needed
Research team member name	

6. **Detailed Summary of evaluation of features:**
 (Overall summary with details on how the requirement/feature is delivering as expected)

7. **Appendix of evaluation of features:**
 (Screenshots & Sources)

Table 5, Internal Testing Template

Finally, we evaluate the collected data and estimate the scoring for each Theme (Table 3).

3.2 Results

The outcome of the evaluation suggests that the system performs well. Most of the requirements were met by the feature sets delivered and tested for each of the cases studies. Table 6 presents a summary of scores for each Theme, analysed by case study and feature.

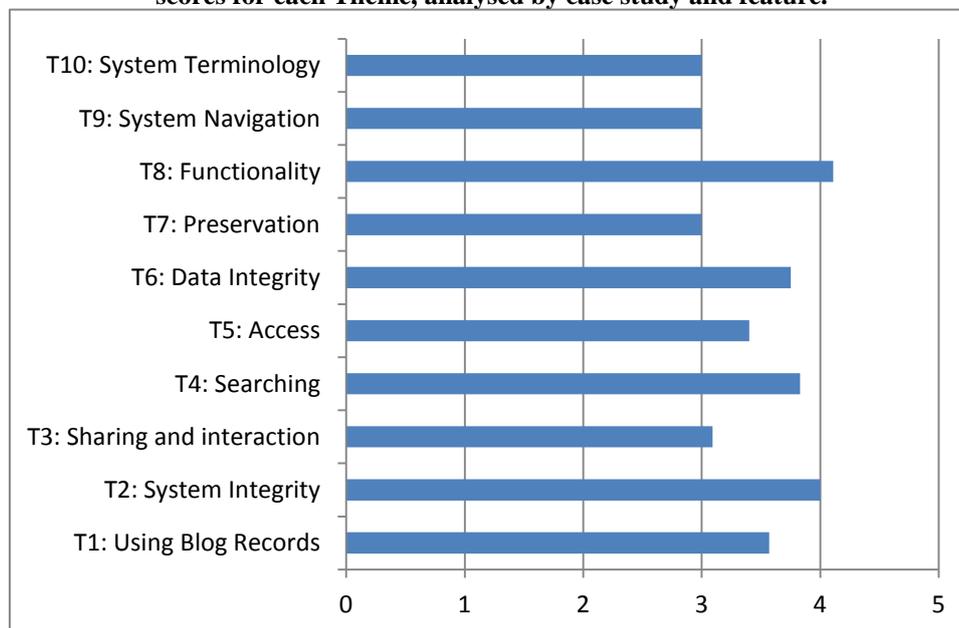


Figure 3 presents a visual representation of the scores for each case study. The detailed information regarding how the internal testing was conducted is explained thoroughly in D5.2 Implementation of Case Studies [1].

	Case Study	Feature	Score	CS score average	Total Theme score average
T1: Using Blog	CS1	RF4	3	3	3.57

Records		RF4	3			
		RF17	4			
		RF6	2			
		CS2	RF23	4	3.8	
			RF25	3		
			RF24	4		
			RF26	4		
			RF28	4		
		CS5	RF66	4	4	
			RF71	4		
			RF72	4		
	CS6	RF65	3	3.5		
		RF83	4			
T2: System Integrity	CS1	RF10	5	5	4.00	
		RF32	N/A			
	CS2	RF18	4	4		
		RF27	4			
	CS6	RF80	3	3.5		
RF79		4				
T3: Sharing and interaction	CS1	RF15	4	4	3.09	
	CS2	RF50	N/A	2.42		
		RF35	2			
		RF31	2			
		RF38	2			
		RF34	3			
		RF47 (old RF 87)	3			
		RF31	3			
	RF31	2				
	CS3	RF31	3	3		
		RF2	4	3.67		
		RF46	4			
	CS4	RF57	2	2		
	CS5	RF38	4	3.5		
		RF46	3			
T4: Searching	CS1	RF69 (old RF68)	4	3.5		
		RF83	3			
	CS3	RF43	N/A	4		
		RF51	4			
	CS5	RF78	4	4		
		RF69	4			
RF73		4				
T5: Access	CS1	RF5	4	3.625	3.4	
		RF5	4			
		RF33	3			
		RF37	N/A			

		RF30	4				
		RF8	4				
		RF13	5				
		RF11	3				
		RF12	4				
		RF19	N/A				
		RF39	4				
		RF84	1				
	CS3	RF62	1	1			
	CS4	RF63	3	3			
RF59		3					
CS5	RF70	4	4				
	RF74	4					
T6: Data Integrity	CS1	SF1	5	4.66	3.75		
		SF7	5				
		RF9	4				
	CS3	RF40	4	4			
		RF41	N/A				
	CS4	RF67	3	3			
		RF56	3				
		RF54	3				
	CS5	RF67	3	3			
	T7: Preservation	CS1	SF16	3		3	3
RF9, RF12, RF18, RF23, RF31, RF40, RF54, RF86, RF87, RF88 and RF89 for DR21: Long term digital preservation requirement			N/A				
CS3		RF87	2	2			
		RF88	2				
CS5		RF40	4	3.33			
		RF87	2				
		RF88	4				
CS6		RF89	4	4			
T8: Functionality		CS1	RF16	5	4.33	4.11	
			RF14	4			
	RF20		4				
	CS2	RF29	4	4			
	CS4	RF14	4	4			
	CS5	RF31	4	4			
		RF86	4				
	CS6	RF82	4	4			
		RF81	4				
	T9: System	CS5	RF1	3	3		3

Navigation					
T10: System Terminology	CS5	RF53	3	3	3

Table 6, Summary of scores for each Theme

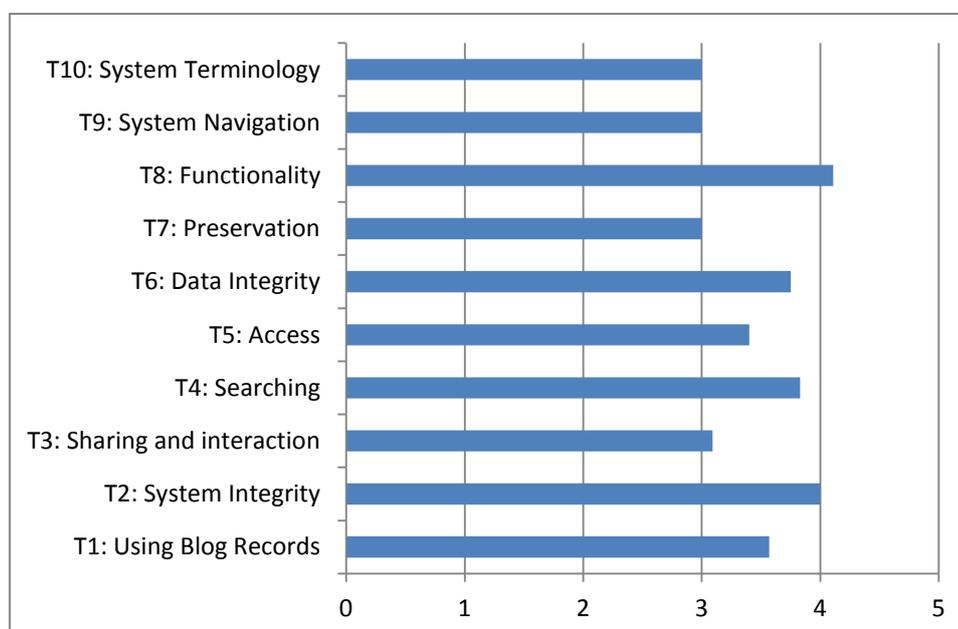


Figure 3, Visual representation of the scores for each case study

In addition to the presented scoring, D5.2 included a summary of the overall results. In each instance, the **expected outcome** from a Research Question (RQ#) is used as a heading:

1. RQ3: Platform is sustainable and it meets the users' needs

I. Relevant results for *system integrity* tests:

- User passwords are being encrypted as required.
- Deduplication (eliminating duplicate copies of repeating data or single-instance storage) and presentation of a single URL for improving referencing performs as expected.
- The repository is capable of handling a large number of user accounts and user activity.

II. Relevant results for *data integrity* tests:

- The spider captures the timestamp and the original URLs.
- Capture of the blog, its comments and its embedded content performs as expected.
- The updated versions of the same content are captured and stored as expected.
- Insufficient contextual metadata about the crawl and the blog was being received from the spider. (This has improved since CS1).
- Content retrieved via APIs is stored in two different databases as part of the preservation strategy.

III. Relevant results for *preservation* tests:

- The evaluation demonstrates that all the data captured by the spider is being ingested into the system; however, improvement is necessary for capturing additional contextual metadata about the crawl and the blog.
- More description is needed with regard to the content of the object (e.g. topic, language, etc.) and to technical aspects (e.g. formats). (This has improved since CS1).
- Content retrieved via APIs is stored in two different databases as part of the preservation strategy.

IV. Relevant results for *functionality* tests:

- The repository provides updates via RSS channel as expected.
- Archive's descriptive statistics and its export operate as expected.
- The platform functions perform as expected to users and administrators.

2. ***RQ4: Searched content is found fast and in an organised manner.***

I. Relevant results for *searching* tests:

- Near-complete search options and features perform as expected, but more salient distinction between record types was recommended.
- External search is also integrated and operates as expected, but additional features for customising are recommended.
- Advanced search included a wide range of metadata options some of which were considered irrelevant.
- The evaluation demonstrates that the system performs well with regards to searching and users' ability to interpret the results.
- The repository is capable of handling a large number of simultaneous searches.

II. Relevant results for *access* tests:

- Some issues related to access to the blog records and representation of dissemination copies of the content were identified.
- The system captures the layout and overall look of blogs as expected.
- The bookmarking, export function and the use of UTF to enable multilingual content operate as expected. However, the translation features perform inconsistently.
- Presented content is harmonised in the repository and is consistently displayed across many different browser types.
- Extraction of content into MARC XML and DC XML performs as expected, but recommendations for including METS/METS XML for import and export are made.
- Extraction as PDF or Image does not perform as expected.
- Support for OpenURL does not perform as expected.
- Navigation of blogs by topics does not perform as expected.
- Improvements are needed for the readability of the statistics graphs.

3. ***RQ5: Usability satisfaction – strengths and weaknesses***

I. Relevant results for *using blog records* tests:

- The evaluation demonstrates that most of the aspects related to the usage of blog records operate as expected.
- Users are able to distinguish between archived and live copies.
- Distinguishing between various types of records was possible, but not intuitive.
- Versioning of records is available, but use of more intuitive terms was deemed necessary.
- Author information and external links are presented as expected.
- The list of blogs is available and navigation through them is possible.
- Further improvements are needed for chronological presentation of blog posts and their tags.
- Improvements were suggested concerning the presentation of posts links.

II. Relevant results for *sharing and interaction* tests:

- The evaluation suggests that content and metadata can be shared with internal and external users through the platform and via social software.
- Dissemination of posts works as expected for various social media services (e.g. Twitter, Facebook), but more sharing options were noted to be desirable.
- Users are able to create personal collections of their favourite blogs and have access to the user history and statistics.
- Citation information and ranking are provided to the user, but improvements in the content as well as the presentation should be made.
- Identifying similar content does not perform as expected.

III. Relevant results for *system terminology* tests:

- The evaluation revealed a number of issues that required clarification of instructions, help pages, and other aspects of terminology in the platform.
- Terminology related to “Submission of Tabs”, “Personalise feature”, “Registration” etc., was often unclear to lay users (i.e. non developers) and required additional information in terms of help pages or change of terms.
- Clarifications in the translations of terms (i.e. Greek) were found necessary. The need for using more specific terms and necessity of additional explanation of the terms was required.

IV. Relevant results for *system navigation* tests:

- The evaluation of general navigational aspects of the platform was performed at various stages of platform’s development. The results suggest the interface to be easy and user-friendly.
- Navigation was generally easy, but improvements were necessary with respect to the prominence of some features (e.g. link to registration, fonts), or system’s feedback messages.
- Navigation through tabs and external links was found useful.

4 External Testing

In this section, we present the External Testing method and results.

4.1 Method

The external users involved in each case study provided their feedback via interviews and associated user questionnaires to gain a better understanding of the functionalities, issues and possible solutions for implementing an improved BlogForever platform. The questionnaire presented in Table 7 is used to gather user feedback.

<p><u>Part A</u></p> <ul style="list-style-type: none"> • A1: First, we would like to understand your background with digital archives or repositories. • A2: Do you actively engage with these digital collections? • A3: Which of the following roles best matches your connection with the discussed digital archive/s? Are you a: Technical Administrator, Content Administrator, Registered User or Guest User? <p><u>Part B</u></p> <ul style="list-style-type: none"> • B1: Complete the registration process. • B2: Submit a blog URL • B3: Using the saving favourites / add to basket function. • B4: Get to the detailed record of a blog • B5: Show citation description within a blog <p><u>Part C</u></p> <ul style="list-style-type: none"> • C1: Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you? • C2: What aspects of the system supported you to perform the exercises or solve the tasks set for you today? • C3: What aspects of the system made it difficult for you to perform the exercises or solve the tasks? • C4: How could the system be improved?
--

Table 7, External Testing Questionnaire

The results of the external evaluation are processed by the authors of this report to gather useful feedback and highlight concrete findings regarding performance and impact. Furthermore, to assist rationalising the external user survey outputs, the questions from parts B and C have been associated with the Themes, as presented in Table 8. In addition to this, each user answer has been assigned a score (range 1 - 5) for each associated Theme. This score is calculated by the authors of this report based on the evaluation of the user's answer (free text).

	T1: Using blog records	T2: System integrity	T3: Sharing and interaction	T4: Searching	T5: Access	T6: Data integrity	T7: Preservation	T8: Functionality	T9: System navigation	T10: System terminology
Part B										
B1: Complete the registration process									X	X
URL								X	X	X
B3: Using the saving favourites/add to basket function			X					X	X	X
B4: Go to the detailed record of a blog	X			X	X	X		X	X	X
B5: Show citation description within a blog	X					X			X	X
Part C										
C1: Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you?		X								X
C2: What aspects of the system supported you to perform the exercises or solve the tasks set for you today?		X								X
C3: What aspects of the system made it difficult for you to perform the exercises or solve the tasks?		X								X
C4: How could the system be improved?		X						X		

Table 8, Association between CS1 external User Questionnaire questions and Research Themes

4.2 Results

The external testing was conducted via selected third party users for each case study. Users submitted a questionnaire containing answers to a variety of questions. These questions were linked to Themes, as presented in Section 4.1. To assist rationalising the users' feedback, questionnaire questions were associated with one or more Themes and a score (range 1-5) was assigned by the authors of the report D5.3 User Questionnaires and Reports. The final scores have been summarised in Table 9.

Themes / Case Studies	CS1	CS2	CS3	CS4	CS5 & CS6	Average
T1: Using blog records	3.57	3.27		4.00	3.30	3.54
T2: System integrity	3.89	3.44	3.25	3.87	3.67	3.62
T3: Sharing and interaction	3.71			4.17	3.23	3.70
T4: Searching	3.43	3.36		3.83	3.20	3.46
T5: Access	3.71	3.36	4.00	3.92	3.60	3.72
T6: Data integrity	3.50	3.50	3.10	4.00	3.70	3.56

T7: Preservation			3.33	4.00	4.00	3.78
T8: Functionality	3.89	3.47	3.50	3.50	3.80	3.63
T9: System navigation	3.54	3.54	3.53	3.87	3.51	3.60
T10: System terminology	3.56	3.63	3.00	3.95	3.43	3.51
Average	3.65	3.45	3.39	3.91	3.54	

Table 9, External Testing Scores Summary

On average, the scores of all case studies are consistent and we consider them to be very positive. The minimum score is 3.45 and the maximum 3.91, which is not such a great variation.

To explore the outcomes even further, we plot the scores grouped by Theme and representing each case study with a different colour (Figure 4). Moreover, we present the average score for each Theme plotted in the range of 4 to 5, in order to highlight more details (Figure 5).

We observe that the lowest scoring Themes are T10: System terminology and T8: Functionality.

On the other hand, the highest scoring Themes are T5: Access and T3: Sharing and interaction.

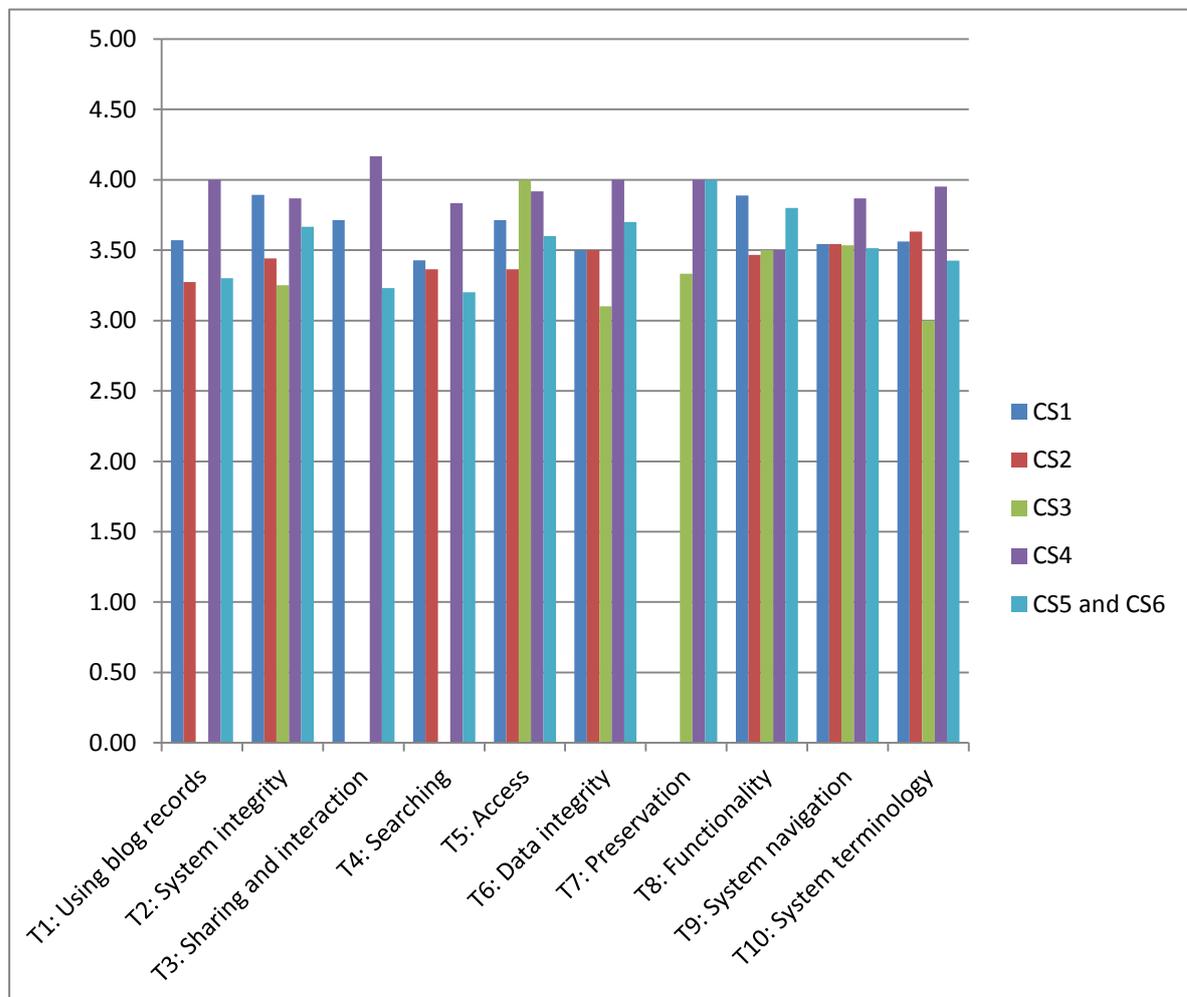


Figure 4, Scores grouped by Theme and Case Study

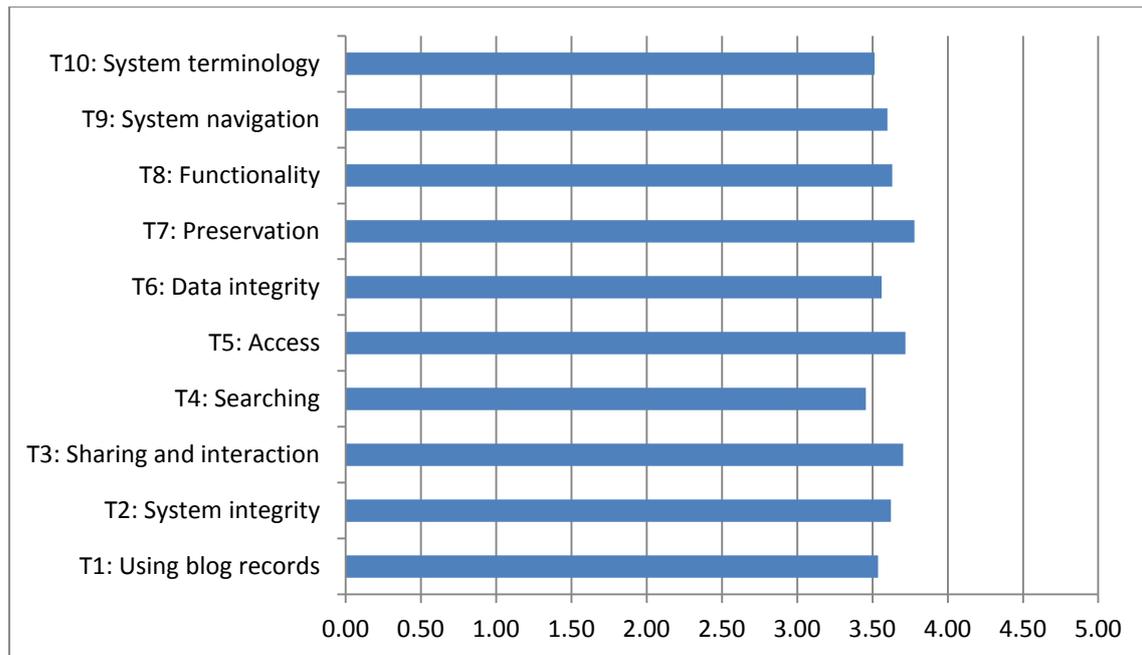


Figure 5, Average scores for each Theme, plotted in range of 4 to 5

Another interesting parameter of the results is the distribution of scores. From Figure 6 to Figure 10, we present the scoring distribution for each case study.

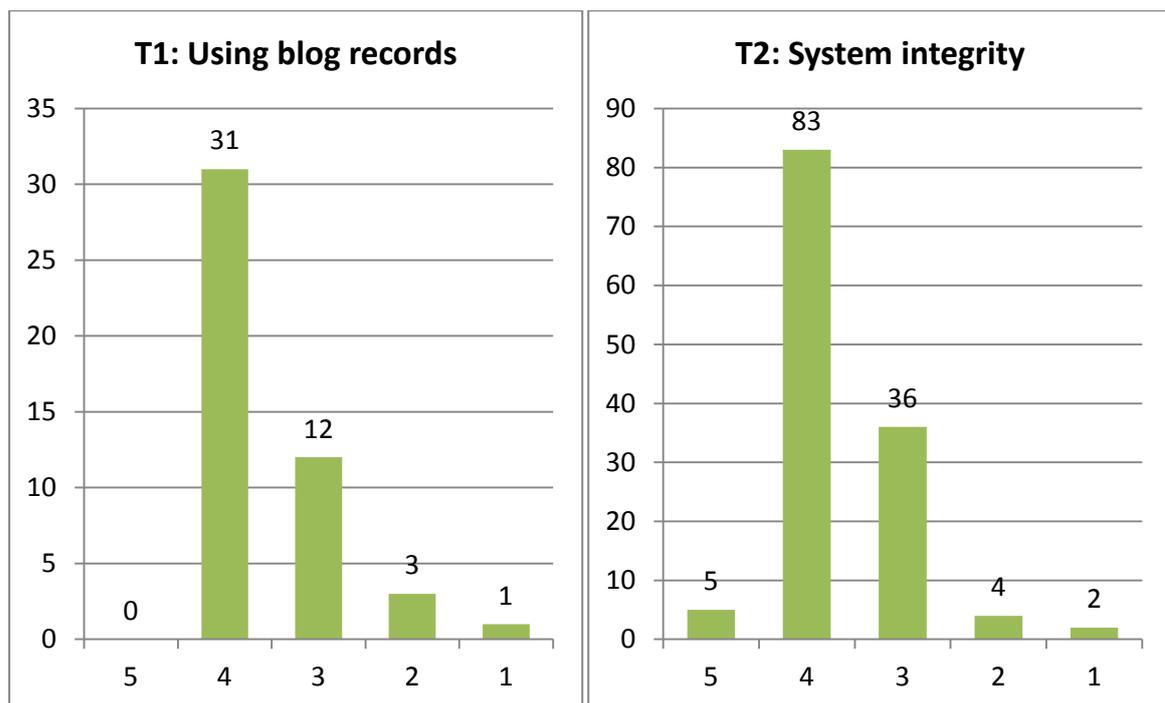


Figure 6, Distribution of scores for T1: Using blog records and T2: System integrity

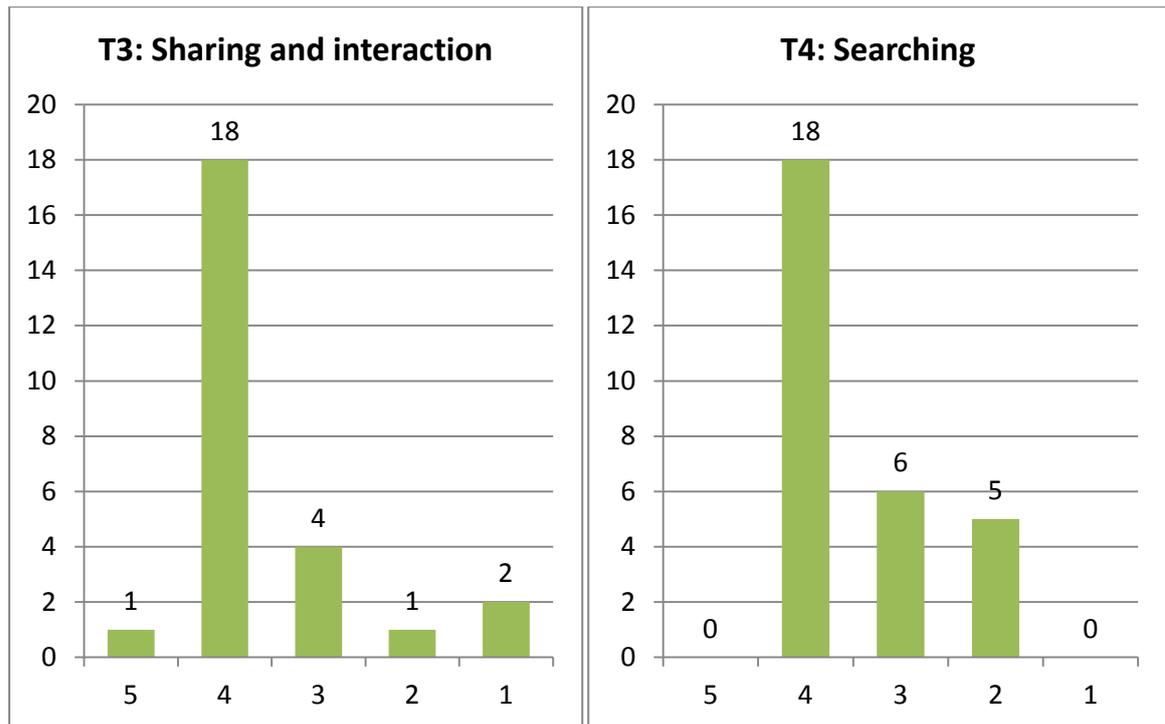


Figure 7, Distribution of scores for T3: Sharing and interaction and T4: Searching

It is impressive most answers scored high in T3: Sharing and interaction. This result indicates beyond all doubt that sharing and interaction is performing very well. On the other hand, T4: Searching also has good scores but there seems to be room for improvement.

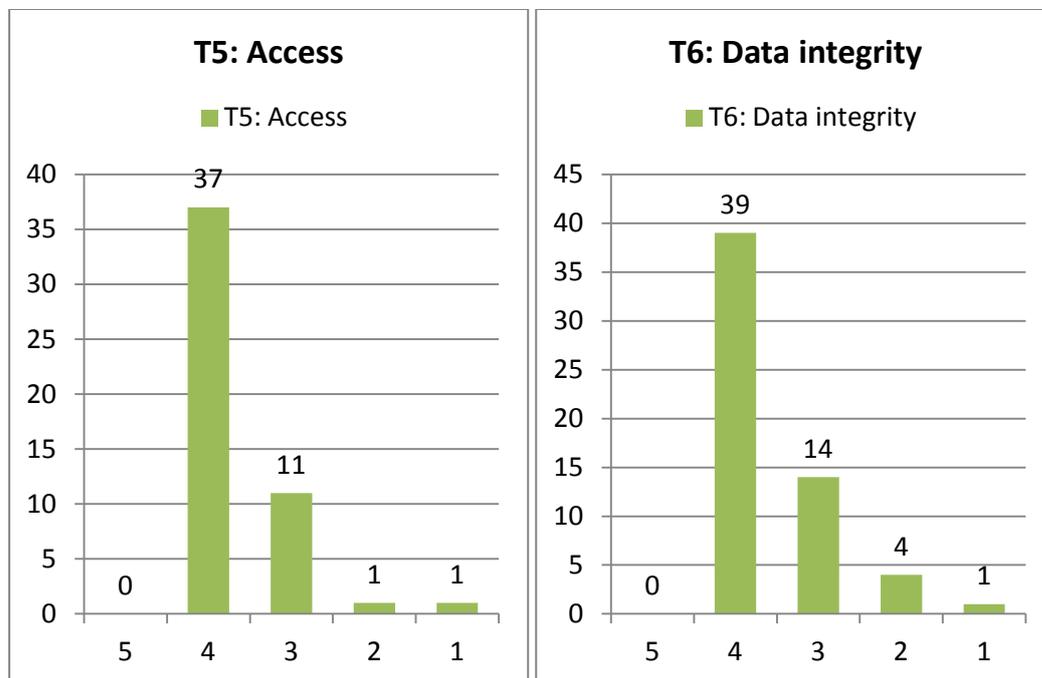


Figure 8, Distribution of scores for T5: Access and T6: Data integrity

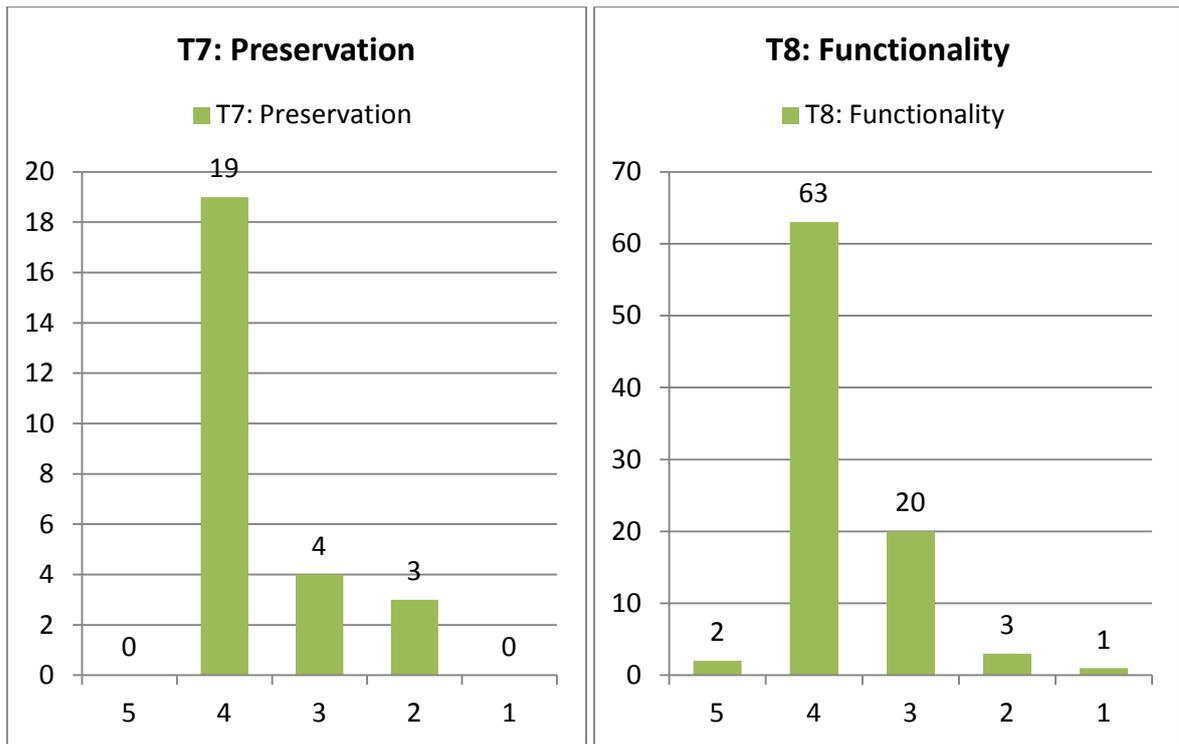


Figure 9, Distribution of scores for T7: Preservation and T8: Functionality

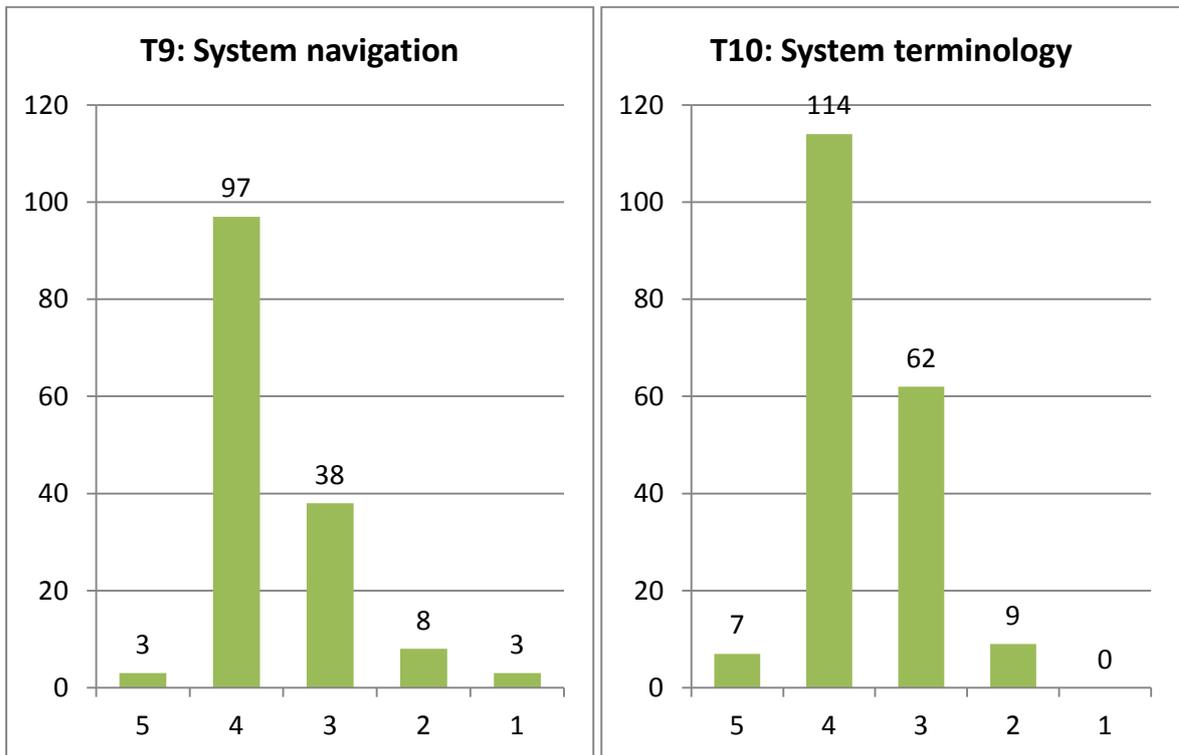


Figure 10, Distribution of scores for T9: System navigation and T10: System terminology

4.2.1 CS1 Answers Summary

This section is a summary of all the answers in user questionnaires.

Direct quotes from users are formatted in italics with inverted commas. Summary evaluations and brief comments from the project team are in standard formatting.

1. **Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you?**
 - *'Pretty good'*
 - *'OK'*
 - *'Fine'*
 - *'Cool!'*
 - *'Good'*
2. **What aspects of the system supported you to perform the exercises or solve the tasks set for you today?**
 - *"Labelling where clear is very useful. It means that as a user you can get to everything you want."*
 - *"Hierarchy of information is very useful. The home page is not cluttered. You don't need to spend a lot of time familiarizing yourself with it."*
 - *"Well structured fairly intuitive."*
 - *"Searching facilities are quite good, well set up, nice and simple. Good to see the narrow by collection option so you can see how many posts are there"*
 - *"Navigation reasonably intuitive, recognisable/familiar layout is extremely important."*
3. **What aspects of the system made it difficult for you to perform the exercises or solve the tasks?**

Registration process noted by all as unintuitive.
Suggested label it as 'login/register'.

"Navigation within the site is not easy. Lack of clarity about how to get to the basket or dashboard. Need more clarity, as some options are not clear, need more labeling and explanations. "

"Unintuitive location of important information (basket, export features) which appears at base of page is not intuitive or easy to find and feels outside the record."
<http://screencast.com/t/y4b7reTeFpL>

People are very used to shopping sites where baskets etc are on top of page e.g. the export functions are too low down.

Not enough **differentiation**. Took some users a while to find detailed record. Not at all clear who **author** is. Need to indicate blog title, blog author. *'Show all'?* *What does this mean?* Not clear where she was. Fixing logic order of title and sub title.

Confusion about inter- connectedness of blogs and posts. of blogs and posts. Are posts searched for from same blog or other blogs. Source blog must always be mentioned in association with post.

Personalise tab gave too many options and confused users, some clearly not relevant such as 'users' and 'loans'.

It seems **special characters** can cause problems in the title at the moment depending on the type of apostrophe (curly or non curly).

Search function, *'very cool!'*

4. How can the system be improved?

All testers reported that finding where to **register** was unclear and most spent one minute locating it at 'login'. Suggest 'login/register' title to this function. Keep in same place. Also indicate that registration is free.

One tester noted *"People who don't use technology can get quite scared and don't need much of an excuse to say bye bye."*

Tabs and breadcrumb trail font is far too small. *"Think about people with visual disabilities. Add 'your notifications' to tab to avoid email glut of notifications"*

Dashboard features lauded but lack of instruction of functionality e.g. *"what happens when you shut down one of the boxes on dashboard? How can you bring it back?..figured it out but guidance needed"*. Finding the **dashboard is not intuitive**, most found it a bit confusing.

Droplist of options from **Personalize tab**, it is clear that some are not relevant, e.g. *"'loans', groups? Please explain or have some help section to expand"*.

Export and basket features very nice and good but location-wise they are too low; they seem like footers and not important information. Adding to personal basket was challenging. Make available at top right hand of the screen, or run alongside from top right hand side of screen. Perhaps call it 'my account' as opposed to 'basket'. Why not allow export to PDF?

Restrictions: Chrome browser didn't give the option to bypass security on laptop but IE did. Chrome worked on her pc. <http://screencast.com/t/6SXF0BUjBi4>. *"Please estate clearly browser requirements"*.

Terminology is unclear to the layperson, i.e. a non developer i.e. *"What does 'similar records' mean?"* Basic understanding/explanation of terms and how to search and what they will find on searching.

"Believe it or not Many researchers/civilians may not understand what a blog is so there should be a simple visual and textual overview of structure of blogs."

Submit tab would be more intuitive if it was labeled 'submit your blog' or words to this effect. Email depositor to notify blog submission is live please.

Search functions and search output: Indentations might help indicate that the posts are coming from blog searched as this is unclear. *"Posts may be coming from different blogs...so we need the title of each blog with the post."*

It is **assumed** the blog/post on top is most recent much in same way a blog works, however this is not stated and it needs to be indicated, so insert date of post as well as source blog of each post in association with each post.

Author should be labeled as such as it is unclear if the name associated with the blog/post/comment is the author. More and improved labeling needed.

Display of search results is unclear, is the default setup to show only three posts? Should state '1st three posts displayed'

Not enough differentiation. Took some users a while to find detailed record. *"Not at all clear who author is. Made no sense to her. Indicate blog title, blog author. 'Show all'? What does this mean?"*
 Not clear where she was. Fixing logic order of title and sub title.

Citation found easily and quickly, however it lacks context about date used, *"what is this date?"*
Must be made clear. Should have date that it was visited in repository by user? Link up with Google scholar?

"Why not have a 'copy/share this' version alongside the citation window. Similar to YouTube function which allows one to copy link and allow sharing."

Sharing feature: It would be good to have a welcome message to repository if people land on site from twitter/facebook etc. Also limit the tabs displayed to people who access the site this way.

It would be good to **grab a screen shot and upload of the original source blog** to preserve look and feel of original as whatever preservation strategy is used, the blog will change in appearance.

4.2.2 CS2 answers summary

CS2 results can be summarised as follows:

1. Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you?

'Pretty good'

'OK'

'Fine'

'Cool!'

'Good'

2. What aspects of the system supported you to perform the exercises or solve the tasks set for you today?

- *"Blog content was well organised."*
- *"Detailed record quite easy to scan, cleanly represented."*
- *"Metadata extracted are accurate (author, date, etc)"*
- *"Good idea to provide citation information."*

3. What aspects of the system made it difficult for you to perform the exercises or solve the tasks?

- Technical issues, bugs submitted
- **Time:** *"Date in the citation information is not clearly defined (is it access date, publication date?)"*
- **Visualisation/Presentation**
 - Images not properly included in the post content
 - Provide more navigation mechanisms
 - *"Browsing the archive is not easy"*
 - List of posts not in reverse chronological order
 - HTML characters not properly processed
 - A lot of vertical scrolling, reduce white space where possible
 - When visiting the blog record, viewing of all posts is not possible
 - Some confusion with terminology
 - Blog repository versus "Blog collection" and "post collection".
- **Content**
 - "Reference links" are not always external (e.g. "jump to comments", links to inline images etc.) or proper (e.g. images should contain link)
 - Export function does not provide adequate metadata

- Consider more citation styles (APA, Harvard etc.)
- No point in providing PDF and JPG export
- Different behaviour for different export links
- BibTex provides information, JPG the actual content
- One case where referential integrity was lost (comment->post)

4. How can the system be improved?

- *"Faceted search is a must!"*
- **Search results:**
 - *"records should be hyperlinked (users clicked randomly either on "HTML" or "detailed record")"*
 - *"Search by URL"*
 - *"Search by date "*
 - *"Some search fields not relevant"*
 - *"Advanced Search form is overloaded with fields."*
 - *"Registration process requires a "veteran" (login page->register form)"*
 - *"Instead of using a JS toggle for citation, provide it as plain text"*
 - *" "Related posts" provided when visiting the post record (not only in search results)."*
- **Date** is crucial metadata. Should be prominent throughout the repository
 - Use timeline for visualising the activity of the blog
- Provide a summary about the blog content (instead of having to actually visit the record)
- Analyse the content (e.g. TagCloud)
- Language filtering (German, French etc.)
- Visual hint about the blog content (e.g. thumbnail)
- Explain some terms (e.g. pages, comments, "collection of" etc.)

4.2.3 CS3 answers summary

Comments:

- Managing a spider without any training is challenging. However, out of the 5 user only one failed in managing the spider based upon the quick guide and a 2min 35sec video with no sound.
- Inserting 10 blogs each, proved close to 90% success rate excluding non-existing blog sites. This is according to the internal, technical and scaled test as well.
- In 20 minutes 180-190 blogs and 10-35 comments was captured – which indicates in average 6 seconds per captured post/comment.
- There is a CSV issue with automated numbering of the inserted blog URL into the default CSV file. This feedback is put on the bug list.
- Feature testing. Most features was found and accepted by all testers. Also some testers had difficulties finding the features in the spider portal. No documented missing features.

Overall impression:

- **Strength:** Speed and user friendly. State of the art especially as it required no training and efficient and not needing manual work load.
- **Weakness:**
 - *"Hard to manage multiple lists of sources on the same spider. "*
 - Difficult finding all output asked for in the portal – most had hard to find number of blogs and comments captured.
 - Some lack of explanation "missing rule", "watchpoint".
- **Improvements:**
 - *"Add a progress bar – for the processing of inserted blog"*
 - Add sound on the instruction video

- Extended status report on the front page
- Graphical of the statistics
- Differentiate the pages in the portal – eg with several colours
- **Bugs:**
 - *"CSV added numbers onto the URLs"*
 - *"Didn't work for Firefox"*

4.2.4 CS4 answers summary

1. **Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you?**
 - *"Pretty good".*
 - *"Excellent"*
 - *"So and so"*
 - *"Good"*

2. **What aspects of the system supported you to perform the exercises or solve the tasks set for you today?**
 - **Tabs** help in navigation: you can see where you are and the adjacent options.
 - **Search interface:**
 - *"only one text box so directly I saw the results, each record and the similar records"*
 - *"interaction was helpful in order to understand what the elements of the system are because of fields search"*
 - **Advanced search**, options like the filters. Some of the users noted that if they had seen the advanced search options from the beginning they would use them.
 - **Links.**
 - **Registration** was an easy task.
 - **Straightforward** way to perform most of the tasks.
 - Pretty **simple** and **straightforward** interface.
 - Green color is nice.

3. **What aspects of the system made it difficult for you to perform the exercises or solve the tasks?**
 - **"Files" tab:** *"Regarding the files and the versions it is not clear what you see. "Files" tab can have a different name that includes the word history or versions."*
 - *"Register" option could be more visible and distinct."*
 - *"Small font-size."*
 - *"Organization is confusing."*
 - *"Words/locales and descriptions."*
 - *"Wrong positions of the components I am looking for."*
 - **Search/Home page:**
 - *"Options under the search input are incomprehensible."*
 - *"It is not clear what the words "collection", "CS3", "CS4" refers to."*
 - *"Small font-size."*
 - *"Green color of the fonts makes them unreadable in combination with the green background in the top of the page."*
 - *"Drop-down menu (any field) show more options in case of the Greek language than in English."*

- *"in a phrase search with 2 words done that there were no results, there was not a clear message that there were no results and the suggestion of results with each of the words confused the user because there was no explanation text."*
- *"I would click on "Search" button but I don't know what "Browse" button does. The results seem different."*
- *"Search results titles aren't related with what I searched."*
- *"Title of a result/post was not linkable."*
- **"Statistics" tab graph:**
 - *"strings are not readable"*
 - *"there is no legend"*
 - *"What the colors represent (purple, blue, grey dotted line)"*
 - *"graph was difficult to understand"*
- **Loss of language:**
 - *"it changes from Greek to English in different times; loss session info."*
- **"Export as"** *"options export different elements and not the same element in a different format."*
- *"English and Greek translations are mixed."*
- *"Basket" (Καλάθι) in Greek is a strange term; it is like a shared repository."*
- *"Usable but with little content."*

4. How can the system be improved?

- *"Bigger tabs and bigger font-size."*
- *"Add border-radius to tabs."*
- *"Links (from Files and Export as) to open in a new tab."*
- **"Export as":**
 - **clearer phrases** (e.g. Export metadata records, Export the reference as BibText) or minimal (as it is) with a **popup** to open on hover with an **explanation** text
 - Could be in a more visible **position**;
 - *"It can be in another Tab because "Export" is important in the preservation concept."*
 - *"Maybe move the element above its current position."*
- *"exported pdf does not include basic information; link to original post, author, if BlogForever generated it then give info about what was the context of the data e.g. when it was crawled"*
- *"Post exported as pdf but didn't contain all the post's data."*
- *"Exported pdf didn't include the images."*
- *"Exported jpeg should not contain html code and should include the post's images."*
- *"Interface can be more user friendly."*
- *"Records don't look like the original posts."*
- *"Colors (green, blue) in my account page (<http://bf3.csd.auth.gr/youraccount/display>) don't match."*
- **In the registration process:**
 - *"Register/login url is not https."*
 - *"after the message that the user received an email with a link to verify his account, a link to register appears again"*
 - *"The message gives the impression that you can login immediately. The message should be clearer that you should see the email in order to proceed."*
 - *"email text: "about 3 days" is not strict, may "about" should be removed."*
 - *"check if username is the same with the password, inform the user is the password is weak/safe"*
 - *"Wrong encoding to the Greek notification letter for the account verification."*
- **Search/Home page:**

- **GUI consistency:** *"blog post search results should have the same format with the blog search results (blog results shows the available files (.jpeg, .txt, xml_mets,) but post results do not contain them)."*
- *"Search results are not clear if they are **ranked** and how they are ranked"*
- *"**Search keywords** could be marked in the results' titles and snippets."*
- *"fonts are not easy to read, maybe a **background** color should exist instead of the white background"*
- **In search results:** *"blog/post title should be link to the record."*
- Snippets should:
 - be representative (the part of the text I am interested in).
 - have the same size (uniform results).
- Encoding problem in the **.htm** pages of **Greek** posts.
- **Words suggestion** after search:
 - words with hyphen in front of them is not clear what they represent
 - html tags should not appear in the suggested words
- Rename **"Detailed records"** to **"show blog"/"show post"**.
- **Logo** antialias.
- Bigger font-size in the **footer**.
 - *"Available languages text is too small."*
- Information **consistency** in the exported formats:
 - *"visually there should not be difference"*
- **"Reference links** (right sidebar) overlap with the content of the post."
- **"Files" tab:**
 - *:Rename it to "Files & Versions" or "Versions" or something that includes the word history."*
 - *"only version one"*
 - *"All the file names have the same date/time. There is **repetition of information** so grouping can be applied."*
- **"Versions** refer to the versions of the post content or to versions of the preservation actions performed:"
 - An explanation text should exist to inform the user what he sees.
- **Keywords** and **references** usually was zero.
 - *"If they are **zero** then the tab to be **inactive**."*
- **"Statistics" tab:**
 - **other posts that other users have seen:**
 - *"there were not enough data in order to see how the list will appear if there were more links (ranking, "see more" option)."*
 - **graph:**
 - *"is incomprehensible"*
 - *"captions are not readable"*
 - *"there is not graph for all the posts, maybe a text "Recorded Downloads: 0" should exist."*
 - *"When 2 different colors are in the same y-axes value, one of the colors should be transparent."*
 - *"Better presentation of the **graph**, maybe 3D and bigger font-size."*
- **Link to the original post** should be more visible:
 - The title of the post in the post page could be link to the original post.
 - Or the link to the **original post** from the bottom of the page can move to the top of it.
- Some texts/strings are not translated in the **Greek** language (e.g. texts inside References, Keywords).

- **Greek** translations like “Δεν έχετε δανειστεί βιβλίο ή δανεισμό” (“You don't have any book on loan.”) need improvement.

4.2.5 CS5 & CS6 answers summary

CS5 and CS6 results can be summarised as follows:

- 1. Please elaborate on how well or badly you feel you performed the exercises/solve the tasks set for you?**
 - *"It took me a few attempts to navigate my around the website. Some tasks were easier than others (easier: submitting a blog; purchasing the restricted collection; more difficult: navigating to a specific blog and highlighting/annotating a section of text; figuring out the digits that appeared once a search for a blog had been performed)"*
 - *"It all seems straightforward enough".*
 - *"Most of it was easy, but the functionality of the timeline was confusing. It was also irritating that there is no dedicated search field for URLs. The purchase of the restricted collection via the “search” button was not intuitive."*
- 2. What aspects of the system supported you to perform the exercises or solve the tasks set for you today?**
 - *"The “Topic” section; tags beneath blog posts".*
 - *"The most helpful things was this information sheet".*
 - *"The navigation bar".*
 - *"Almost every time I used the search. The search was very helpful".*
- 3. What aspects of the system made it difficult for you to perform the exercises or solve the tasks?**
 - *"The help central needs to be extended".*
 - *"I had a few difficulties returning to the dashboard once I began my searches".*
 - *"I think things generally need to be a bit clearer. But then, perhaps that comes with use".*
 - *"Sometimes the confusing interfaces (see above). Short help hints e. g. in the submit section would help".*
- 4. How can the system be improved?**
 - *"The search functionality needs to be improved".*
 - *"Some instructions on the “Home” section on how to use the website, apart from having them hidden in the Help section would be beneficial for first-time users".*
 - *"I also noticed that sections of the blogs overlapped other sections. Perhaps they need to be resized or reformatted".*
 - *"Maybe a clearer separation of the action related to searching and to add/change something. It is not obvious that some of the entries in the navigation bar are just filters for global search".*
 - *"Some graphical things would help to focus. On the left I have many meta data options and stuff and it is difficult to tell what I am looking for. It attracts my attention away from the blogpost".*

The combined conclusions of all evaluation methods are presented in Section 6, Conclusions.

5 System Logs

In this section we present the method and outcomes of the BlogForever System Logs monitoring.

5.1 Method

The evaluation is based on the observation of the user behaviour and the tracking of user actions throughout the BlogForever instances deployed for the case studies. The data gathered from System logs and analysed to specify Metrics relevant to Research Questions and Themes. The defined Metrics are presented in Table 10 and their association with Themes is presented in Table 11.

ID	Metric	Description
M1	Content records page views	Repository pages presenting records (blogs, posts, pages or comments). Example: http://bf3.csd.auth.gr/record/5394?ln=en
M2	Export page views	Repository pages used to export content. Example of record export: http://bf3.csd.auth.gr/record/5394/export/hx?ln=en Example of search results export: http://bf3.csd.auth.gr/search?ln=en&p=document&f=&action_search=Search&c=BlogForever+Test+Repository&sf=&so=d&rm=&rg=10&sc=1&of=xm
M3	Search page views	Repository pages used for search. Example: http://bf3.csd.auth.gr/search?ln=en&sc=1&p=analysis&f=&action_search=Search&c=Blogs&c=Posts&c=Comments&c=Pages
M4	Achieve goals in Google Analytics	Goals are a versatile way to measure how well a website fulfils specific objectives, which can be a set of consecutive actions in the website.
M5	Number of python code errors	The number and nature of python errors is important for the system integrity.
M6	HTTP status distribution	The distribution of HTTP responses provides an insight on application stability and integrity.
M7	Page loading time distribution	Average web page loading time is a characteristic of website performance.
M8	Pages per visit	The more pages per user visit is directly relevant to the quality of system navigation and functionality.
M9	Average visit duration	The length of user visit duration is directly relevant to the quality of system navigation and functionality.

Table 10, Metrics [15]

Metrics / Themes	T1: Using blog records	T2: System integrity	T3: Sharing and interaction	T4: Searching	T5: Access	T6: Data integrity	T7: Preservation	T8: Functionality	T9: System navigation	T10: System terminology
M1: Content records page views	X				X					
M2: Export page views			X		X					
M3: Search page views				X						
M4: Achieve goals in Google Analytics								X	X	X
M5: Number of python code errors		X						X		
M6: HTTP Status Distribution		X						X		
M7: Page loading time distribution		X								
M8: Pages per visit								X	X	
M9: Average visit duration								X	X	

Table 11, Metrics and Themes association [10]

As we already discussed earlier in Section 2.2 Research Questions and Themes, the Themes are only linked to RQ3, RQ4 and RQ5. In addition, System Logs analysis is capable of answering also RQ1 and RQ2.

RQ1 is defined as follows:

*What are the particular **problems** the implementation is facing? Or are the BlogForever software implementation processes an **overall success**?*

The Metrics related to system integrity (M5: Number of python code errors, M6: HTTP Status Distribution, M7: Page loading time distribution) are directly relevant to RQ1.

RQ2 is defined as follows:

*Are complex BlogForever platform search strategies **working efficiently** when high levels of content are available within the BlogForever platform?*

System Logs analysis will be capable of answering RQ2 by analysing the data of the biggest cases studies which will involve large data sets.

5.2 CS1 & CS2 Results

The results from the first two case studies present the status of the BlogForever platform in a positive light. Detailed data was recorded regarding visitors, page views and other major statistics, strengthening the support of our evaluation. The overall usage data are presented in Table 12.

	<i>Sep</i>	<i>Oct</i>	<i>Nov</i>	<i>Dec</i>
Unique Visits	84	85	67	24
Number of Page views in the Platform	1,251	1,387	1,250	62
Average Page/Visit	14.89	16.32	18.66	2.58
Average time on site	14:55	12:09	11:15	01:48

Table 12, Overall usage data for CS1 & CS2 [10]

CS1 and CS2 have been visited 240 times (), an adequate number of visitors to draw conclusions. Furthermore, it is also interesting to note that 3.950 page views have been performed, bringing the analogy of pages per visit to 15.19.

The Metrics presented here are related to the Topics. Each Metric is linked to one or more Topics which in turn are linked to Research Questions. Using Metrics, we are able to isolate specific system log variables which are relevant to the research we are performing during the BlogForever Case Studies.

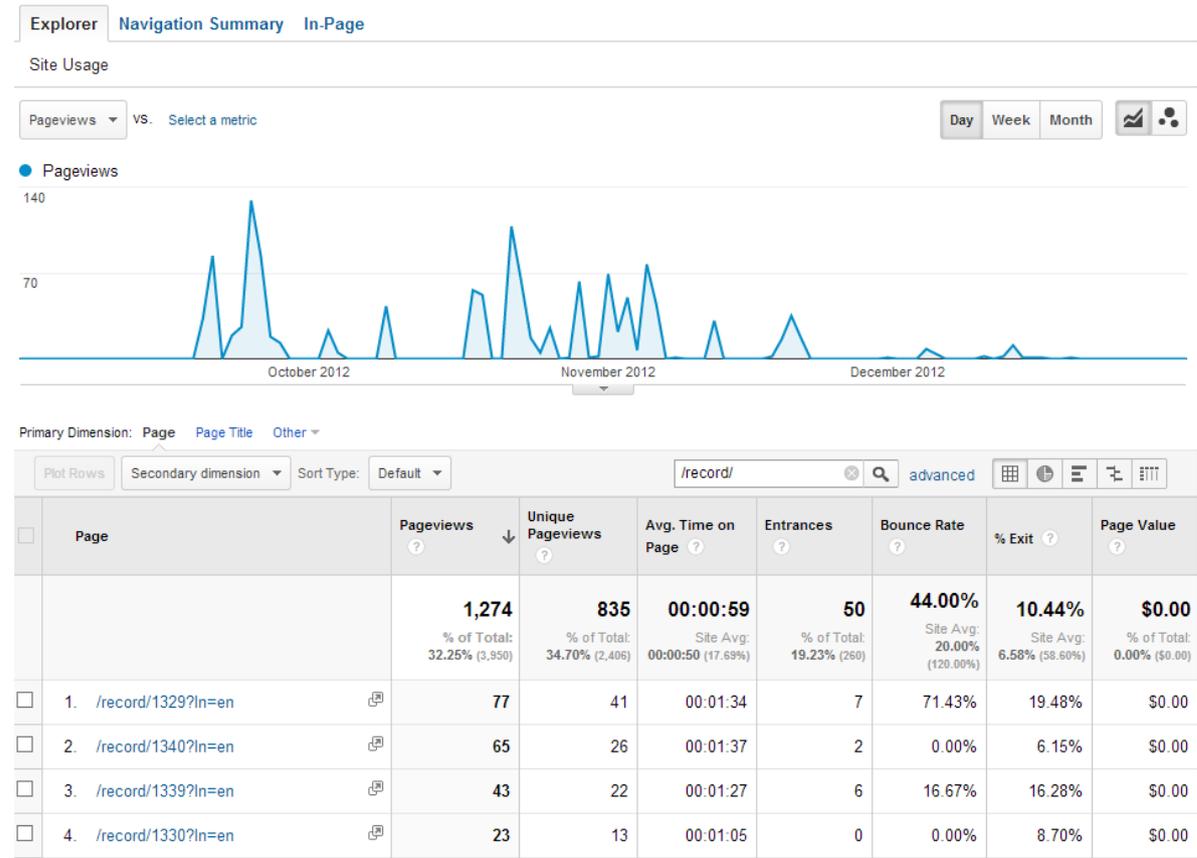


Figure 11, CS1-CS2 Metric 1: Content records page views

Content records page views is an important Metric as the content records are the parts of the repository which actually contain the information that the user seeks. In Figure 11, we can see that the content records page views for CS1-CS2 are 32.25% of the total page views. This outcome reveals that the repository content was very accessible to the users.

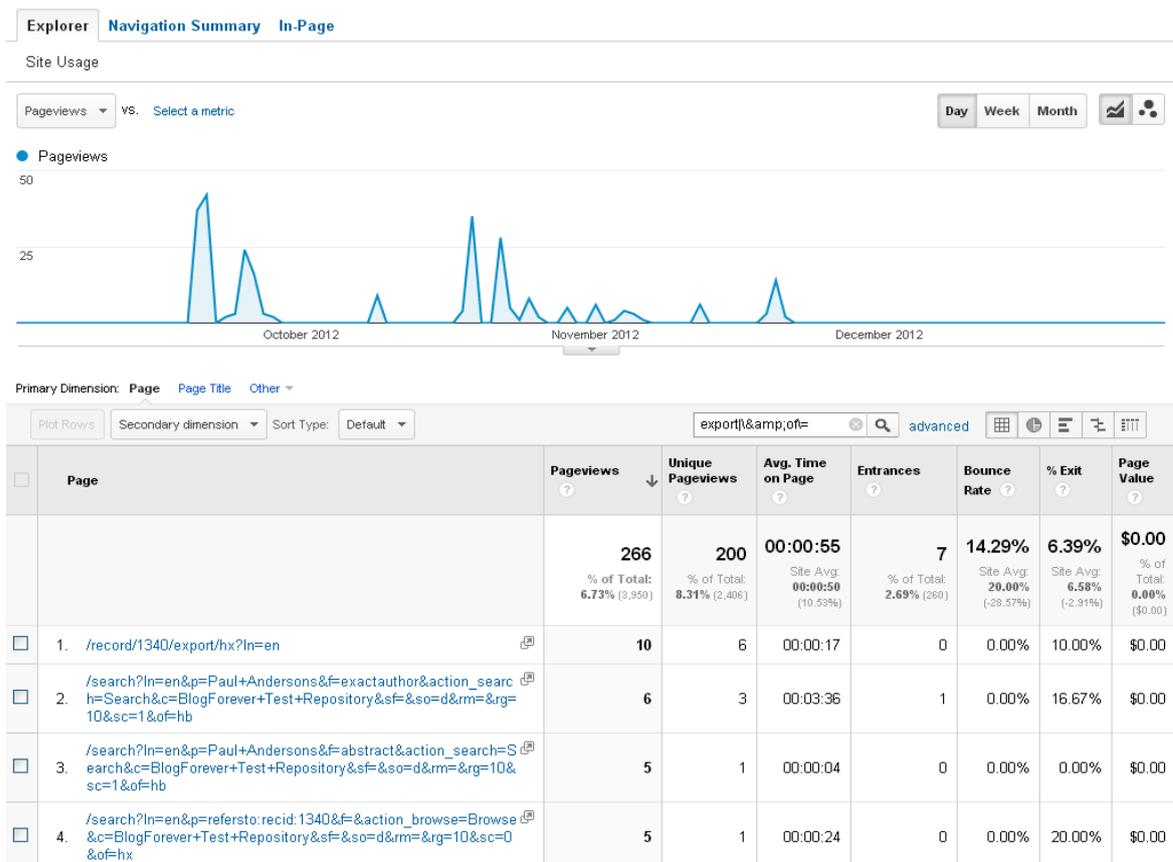


Figure 12, CS1-CS2 Metric 2: Export page views

Export page views were also accessed (6.73% of all requests), as presented in Figure 12. This number is quite low compared to the record page views but this is normal because only a single GA Goal involving exporting content was defined in the case studies.

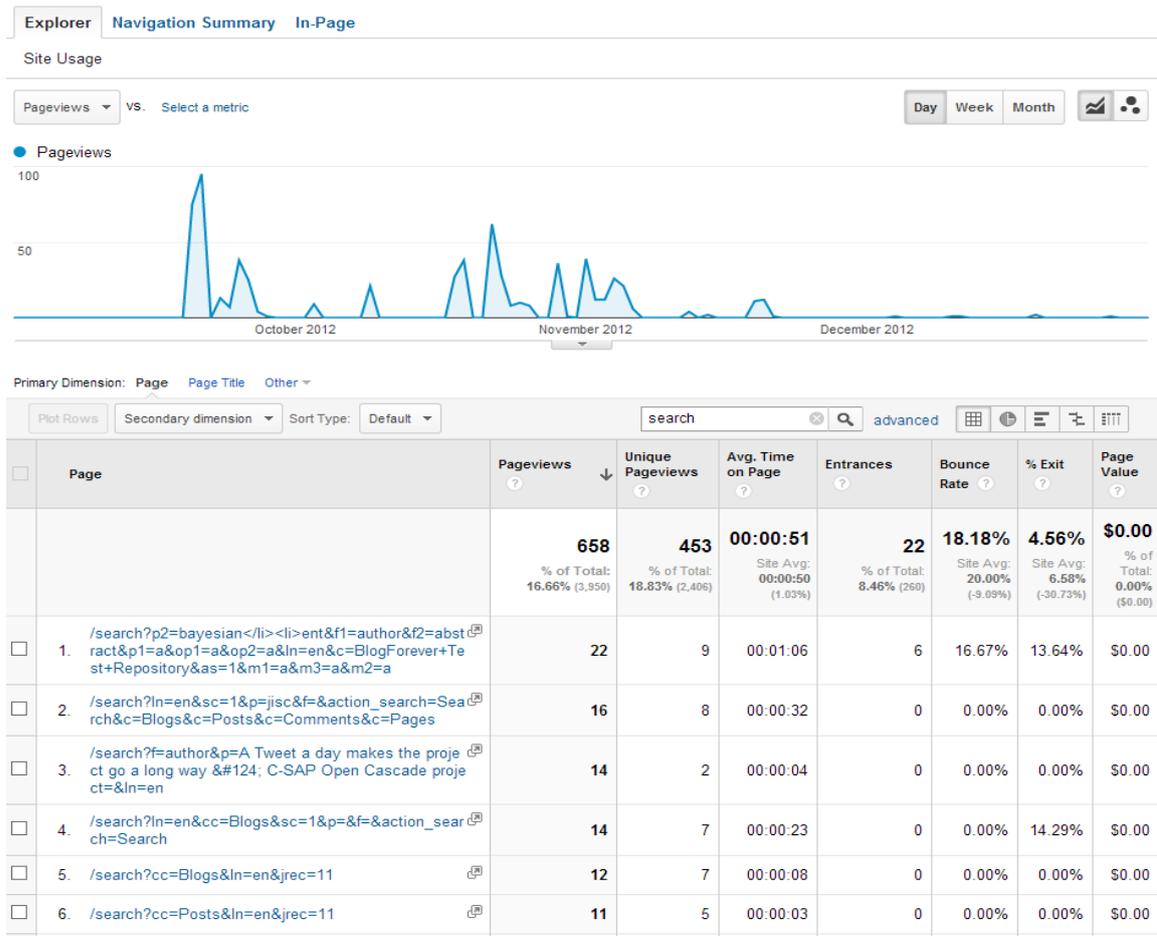


Figure 13, CS1-CS2 Metric 3: Search page views.

Figure 13 presents the search page views. An interesting fact is that the exit percentage is 4.56%. This means that only 4.56% of users exited the repository after performing a search. The vast majority continued navigation, implying that the search results were satisfactory. To compare this with previous results, exit percentage in content record page views (Figure 11) was 10.44%, which is more than double.

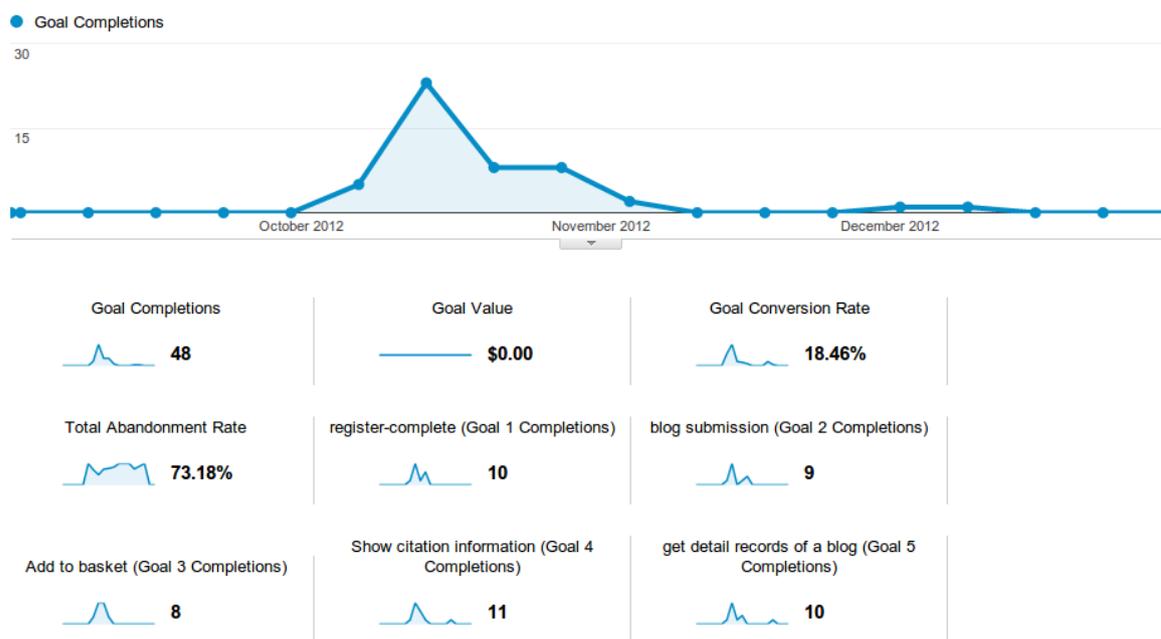


Figure 14, CS1-CS2 Metric 4: Achieved goals in GA.

Figure 14 presents achieved goals in GA. The total abandonment rate of 73,18% seems problematic. This indicates functionality and navigation issues for the platform as the Goals set for users could not be achieved.

<i>Number</i>	<i>Error</i>
26352	traceback.print_stack()
23939	if not self._defer_warnings
23620	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 168, in execute
23620	warn(w[-1], self.Warning, 3)
23620	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 82, in _warning_check
23619	File "/usr/local/lib/python2.6/dist-packages/invenio/webinterface_handler_wsgi.py", line 462, in application
23595	rc = cur.execute(sql, param)
23595	result = _check_result(req, obj(req, form))
23595	return root._traverse(req, path, False, guest_p)
23595	return _handler(req)

Table 13, CS1-CS2 Metric 5: Number of python code errors

A further analysis of the python error log by invenio module name is also possible because by convention, each invenio source code file name has a prefix equal to the module it belongs to. Thus, we checked for all invenio module names (*bibauthorid*, *bibcatalog*, *bibcheck*, *bibcirculation*, *bibclassify*, *bibconvert*, *bibdocfile*, *bibedit*, *bibencode*, *bibexport*, *bibformat*, *bibindex*, *bibknowledge*, *bibmatch*, *bibmerge*, *bibrank*, *bibrecord*, *bibsched*, *bibsort*, *bibsword*, *bibupload*,

blogspam, docextract, elmsubmit, oaiharvest, oairepository, refextract, utils, webaccess, webalert, webauthorprofile, webbasket, webdeposit, webhelp, webjournal, weblinkback, webmessage, websearch, websession, webstyle, websubmit) as well as generic python module names (*pyPdf, errorlib.py, sets.py, MySQLdb*). The outcomes of this analysis reveal the more problematic parts of the source code:

<i>Module</i>	<i>Errors</i>
MySQLdb	47569
invenio errorlib	29134
invenio websearch	26225
invenio bibrank	23471
inveniowebsubmit	8477
pyPdf	5514
sets.py	5514
invenio bibdocfile	2705
invenio utils	492
invenio websession	125
invenio webstyle	8

Table 14, CS1-CS2 Metric 5: Errors grouped by python module

<i>Number</i>	<i>Percent</i>	<i>Status Code</i>	<i>Explanation</i>
837272	83.52%	200	OK - The request sent by the client was successful
131294	13.10%	404	Document Not Found - Requested resource could not be found
32721	3.26%	302	Moved Temporarily (redirect)
985	0.10%	304	Not Modified - Resource has not been modified
227	0.02%	500	Internal Server Error
12	~ 0%	403	Forbidden - Server is refusing to respond to it

Table 15, CS1-CS2 Metric 6: HTTP status distribution

A further examination of the source of the erroneous status codes reveals the following information:

- 128,063 not found errors (404) are attributed to a missing robots.txt file, a trivial issue affecting web bots visiting the website. This issue is irrelevant to the case studies.
- 1,136 not found errors (404) are attributed to a missing /favicon.ico file. Browsers are using favicons images to display a website logo in their address bar. Thus, this issue is also irrelevant.

- The remaining 2,095 not found errors (404) are actual issues of missing web resources. This number is extremely low and can be considered normal.
- The redirect (302) and not modified (304) status codes are issued by the software and do not constitute a problem.
- The internal server errors (500) are caused by actual problems in the platform. Their number is extremely low (0.02% of the total number of requests).

To conclude with Metric 6: HTTP status distribution, it is evident that the BlogForever platform is working very well and this is indicated by the log file statistics which have practically no errors while the not found HTTP requests are extremely low (2,095, 0.002%).

Page Load Time Bucket (sec)	Page Load Sample	Percentage of total
0 - 1	64	14.00% 
1 - 3	223	48.80% 
3 - 7	110	24.07% 
7 - 13	37	8.10% 
13 - 21	11	2.41% 
21 - 35	5	1.09% 
35 - 60	6	1.31% 
60+	1	0.22%

Table 16, CS1-CS2 Metric 7 Page loading time distribution

Table 16 presents the page loading time distribution. The vast majority of pages (~90%) load in less than 7 seconds. The performance of the repository is considered very satisfactory according to current web standards.

Pages / Visit	15.19
---------------	-------

Table 17, CS1-CS2 Metric 8: Pages per visit

Average Visit Duration	11:51
------------------------	-------

Table 18, CS1-CS2 Metric 9: Average Visit Duration

Finally, Tables Table 17 and Table 18 present some data on the average pages per visit and the average visit duration. This information highlights the fact that the users spent considerable time in the repository and conducted thorough testing.

To conclude, the first two BlogForever Case Studies provided us with useful feedback. A considerable number of users were involved and many different features were tested. The results show that the platform is robust and does not suffer from many errors. Performance is also good. Navigation and search seem to work well. The only issue discovered has to do with the very low achievements of Goals, which is attributed to expected limited functionality and navigation issues, as these are the first case studies.

5.3 CS3 & CS4 Results

The results from the second instance of the BlogForever platform are also quite satisfactory.

CS3 and CS4 general information is presented in Figure 15 and Table 19. They have been visited 958 times, considerably more than CS1 & CS2 (240). Also 8,330 page views have been performed and the average number of pages per visit is 8.7.

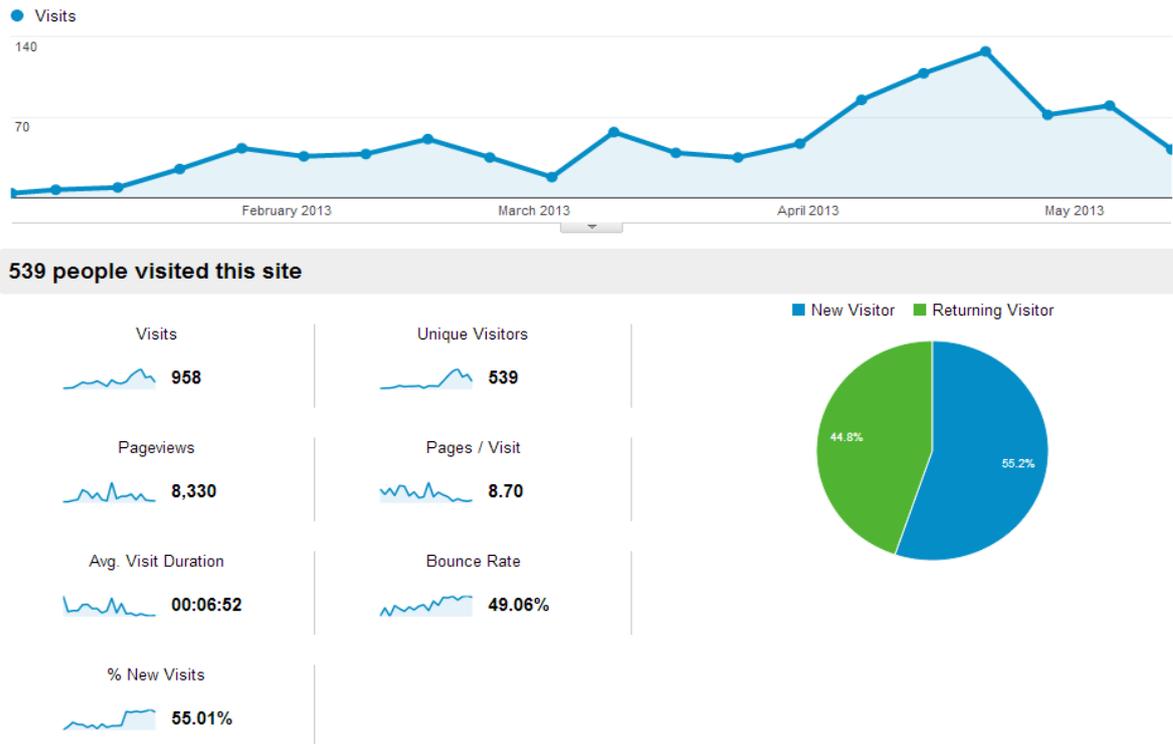


Figure 15, General information on visits for CS3 & CS4

	Nov	Dec	Jan	Feb	Mar	Apr	May
Unique Visits	15	55	85	153	161	407	172
Number of Page views in the Platform	87	582	1,472	1,940	2,477	2,116	353
Average Pages/Visit	5.8	10.58	17.32	12.68	15.39	5.20	2.05
Average time on site	05:41	05:31	13:17	11:05	15:25	02:53	00:43

Table 19, Overall usage data for CS3 & CS4.

Using Metrics, we are able to isolate specific system log variables which are relevant to the research we are performing during the BlogForever Case Studies.

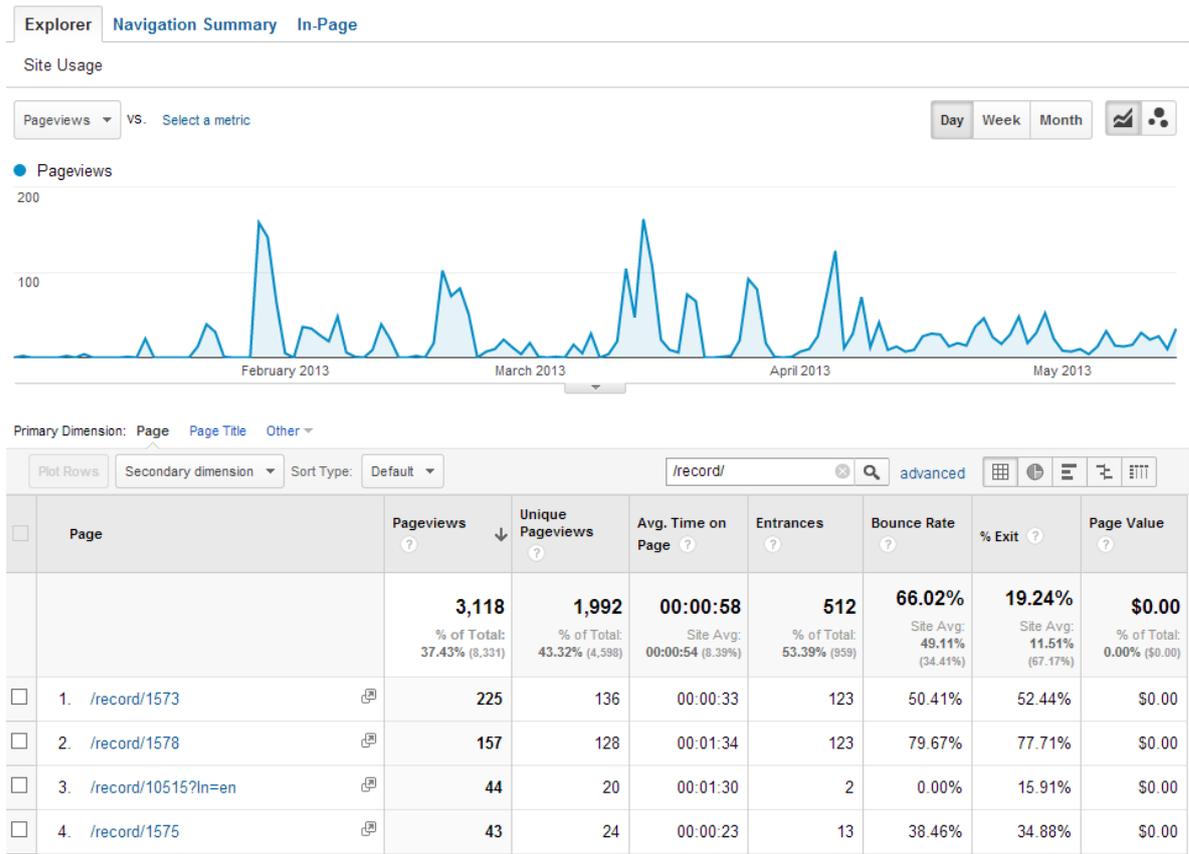


Figure 16, CS3-CS4 Metric 1: Content records page views

Content records page views is an important Metric as the content records are the parts of the repository which actually contain the information that the user seeks. In Figure 16, we can see that the content records page views for CS3-CS4 are 37.43% of the total page views. Over one third of the page views are content records views, indicating that content is easily reachable from users.

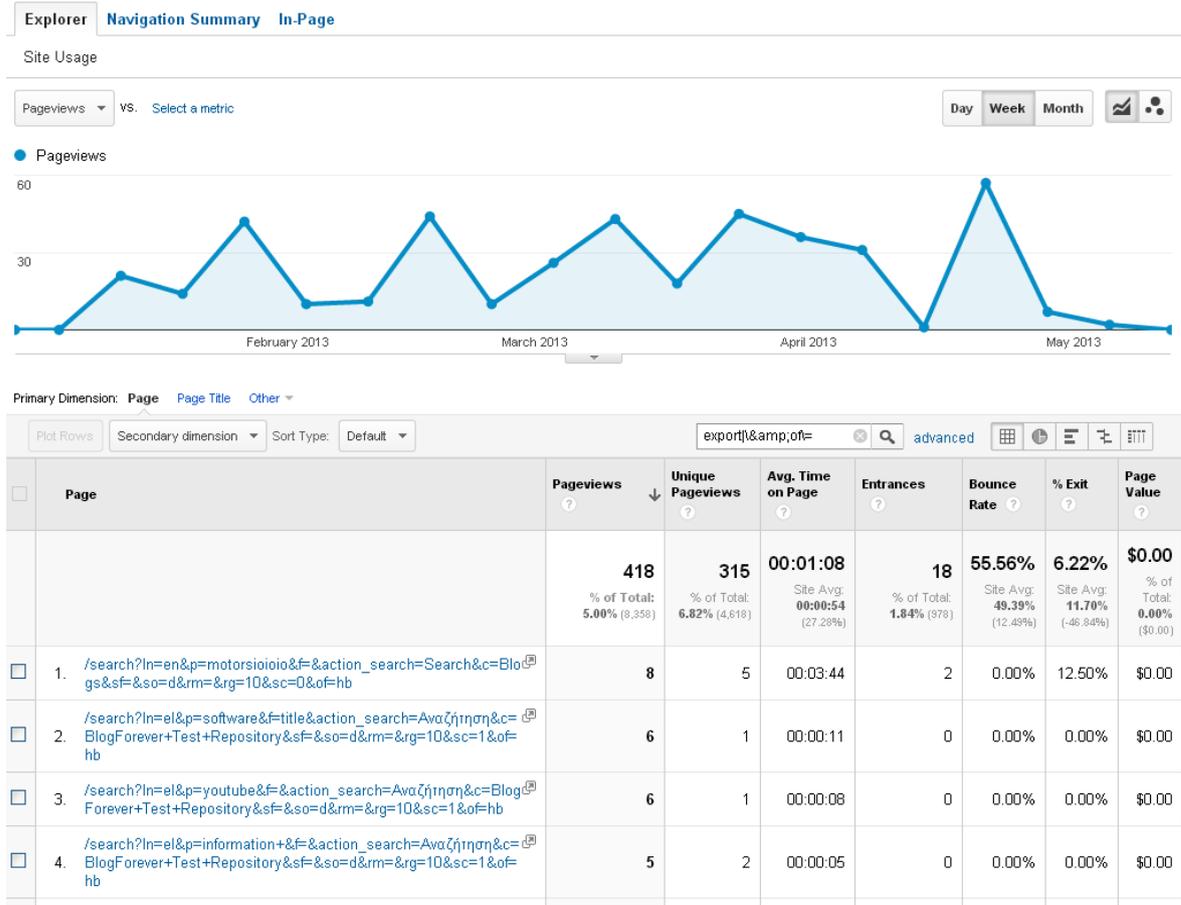


Figure 17, CS3-CS4 Metric 2, Export page views

Export page views were also accessed (5% of all requests), as presented in Figure 12. This number is quite low, even lower than CS1 & CS2 where it was 6.73%.

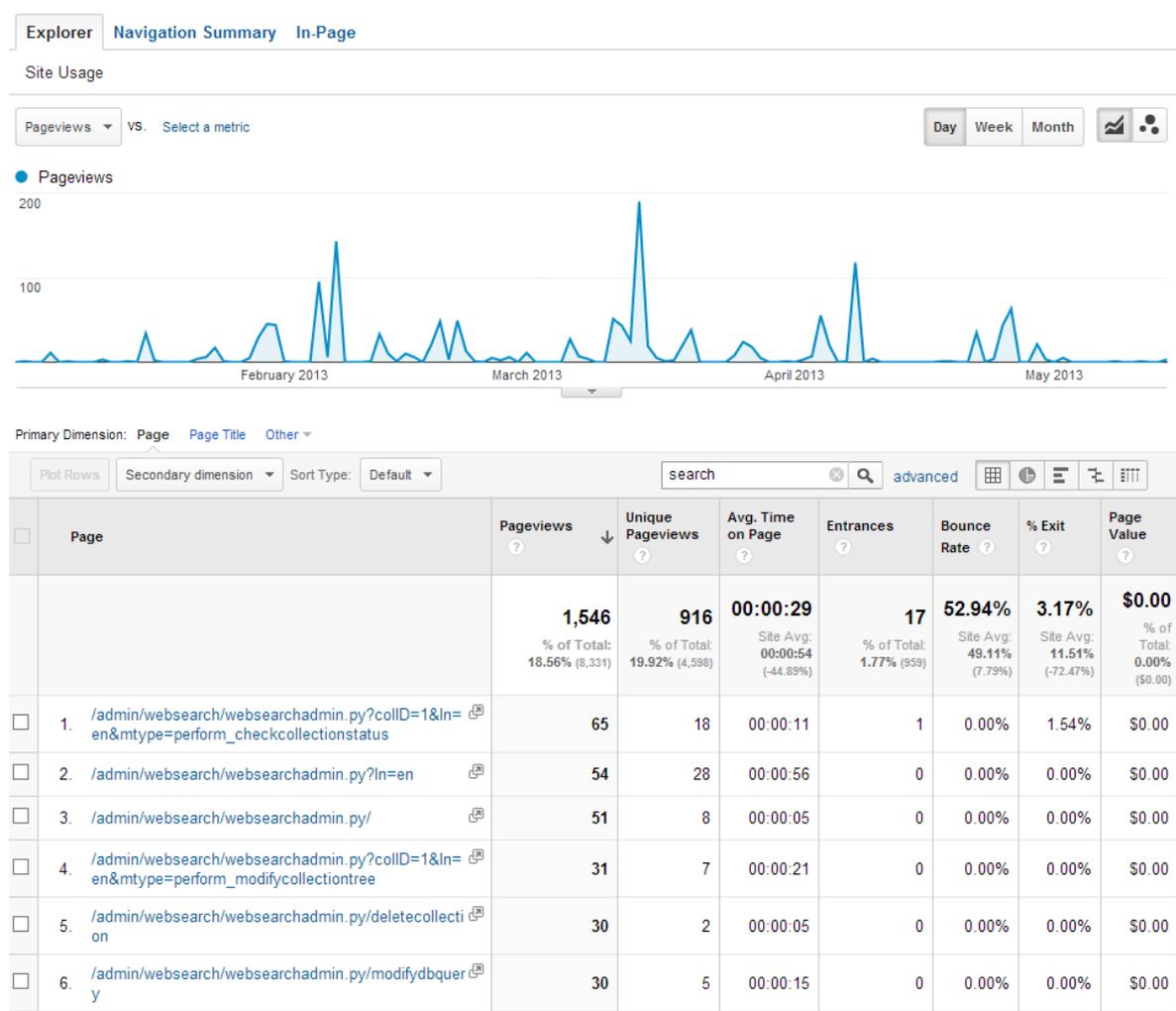


Figure 18, CS3-CS4 Metric 3: Search page views

Figure 18 presents the search page views (18.56%) which have increased slightly compared to CS1-CS2 (16.66%). The exit percentage is 3.17%, lower than the CS1-CS2 exit percentage of 4.56%. This is an indication that search has improved since the first two case studies.

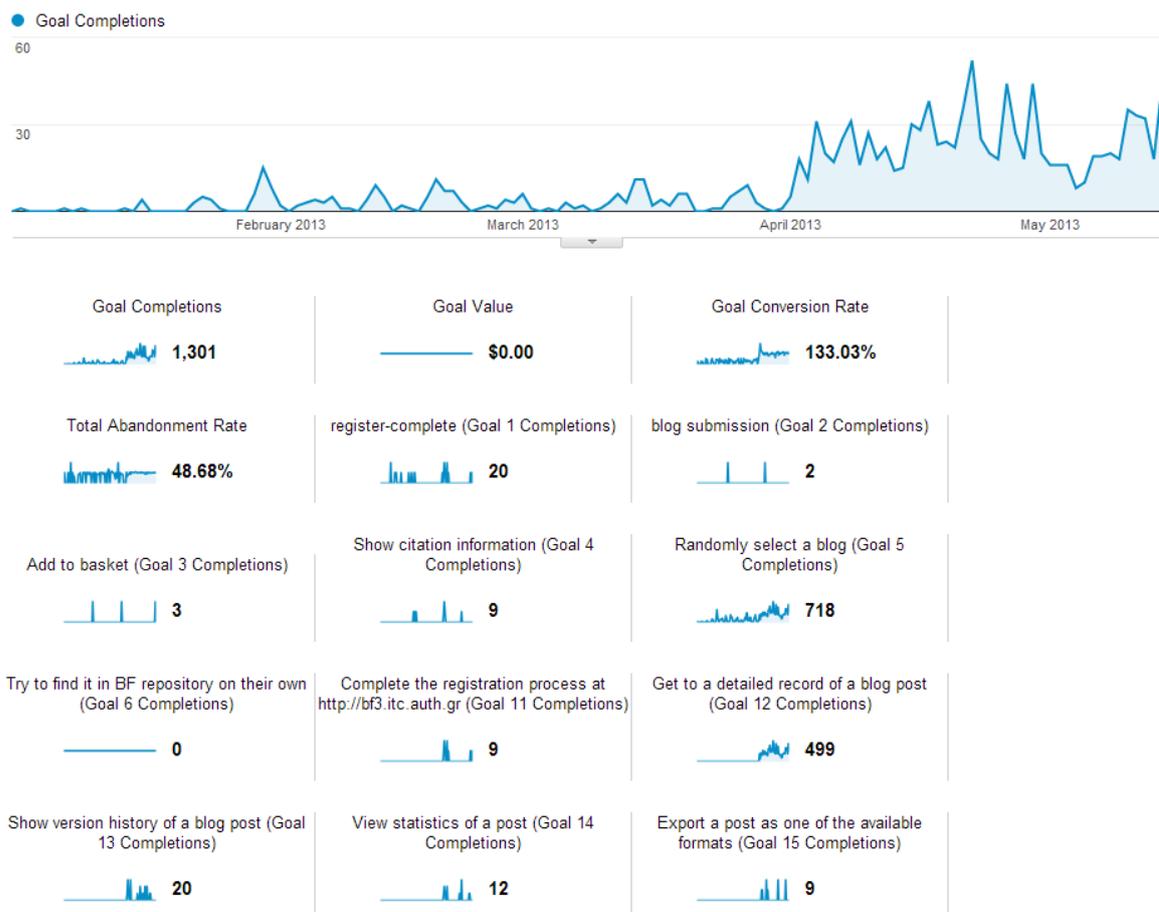


Figure 19, CS3-CS4 Metric 4: Achieved goals in GA

Figure 19 presents the achieved goals in GA, where we see a great improvement since the previous case studies. The Total Abandonment Rate is 48.68%, whereas in CS1-CS2 it was 73.18%. This finding indicates that the functionality and navigation has improved, even though it is still quite low. Approximately half of the visitors cannot complete their goals.

<i>Number</i>	<i>Error</i>
14630	traceback.print_stack()
14630	if not self._defer_warnings
14600	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 168, in execute
12550	warn(w[-1], self.Warning, 3)
12550	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 82, in _warning_check
12457	File "/usr/local/lib/python2.6/dist-packages/invenio/webinterface_handler_wsgi.py", line 462, in application
12443	rc = cur.execute(sql, param)
12443	result = _check_result(req, obj(req, form))
12443	return root._traverse(req, path, False, guest_p)
12443	return _handler(req)

Table 20, CS3-CS4 Metric 5: Python code errors

A further analysis of the python error log by invenio module name is also possible because by convention, each invenio source code file name has a prefix equal to the module it belongs to. Thus, we checked for all invenio module names (*bibauthorid*, *bibcatalog*, *bibcheck*, *bibcirculation*, *bibclassify*, *bibconvert*, *bibdocfile*, *bibedit*, *bibencode*, *bibexport*, *bibformat*, *bibindex*, *bibknowledge*, *bibmatch*, *bibmerge*, *bibrank*, *bibrecord*, *bibsched*, *bibsort*, *bibsword*, *bibupload*, *blogspam*, *docextract*, *elmsubmit*, *oaiharvest*, *oairepository*, *refextract*, *utils*, *webaccess*, *webalert*, *webauthorprofile*, *webbasket*, *webdeposit*, *webhelp*, *webjournal*, *weblinkback*, *webmessage*, *websearch*, *websession*, *webstyle*, *websubmit*) as well as generic python module names (*pyPdf*, *errorlib.py*, *sets.py*, *MySQLdb*). The outcomes of this analysis reveal the more problematic parts of the source code:

<i>Module</i>	<i>Errors</i>
MySQLdb	79289
invenio errorlib	29724
invenio bibrank	21152
invenio websearch	21150
invenio bibdocfile	16726
invenio utils	946
websubmit	70
bibformat	23
invenio websession	16
bibknowledge	12
invenio webstyle	11
bibcirculation	7

Table 21, CS3-CS4 Metric 5: Errors grouped by python module

<i>Number</i>	<i>Percent</i>	<i>Status Code</i>	<i>Explanation</i>
512,620	82.81%	200	OK - The request sent by the client was successful
77,831	12.57%	404	Document Not Found - Requested resource could not be found
18,268	2.95%	302	Moved Temporarily (redirect)
10,041	1.62%	304	Not Modified - Resource has not been modified
152	0.02%	500	Internal Server Error
73	~0%	403	Forbidden - Server is refusing to respond to it

Table 22, CS3-CS4 Metric 6: HTTP Status distribution

A further examination of the source of the erroneous status codes reveals the following information:

- 55,545 not found errors (404) are attributed to a missing robots.txt file, a trivial issue affecting web bots visiting the website. This issue is irrelevant to the case studies.
- 10,008 not found errors (404) are attributed to a missing /favicon.ico file. Browsers are using favicon images to display a website logo in their address bar. Thus, this issue is also irrelevant.
- The remaining 12,278 not found errors (404) are actual issues of missing web resources. This number is extremely low and can be considered normal.
- The redirect (302) and not modified (304) status codes are issued by the software and do not constitute a problem.
- The internal server errors (500) are caused by actual problems in the platform. Their number is practically zero.

To conclude on Metric 6: HTTP status distribution, it is evident that the BlogForever platform is working very well and this is proven by the log file statistics which have practically no errors while the not found resources are only 12,278 (1.9% of all HTTP requests).

Page Load Time Bucket (sec)	Page Load Sample	Percentage of total
0 - 1	45	9.09%
1 - 3	208	42.02%
3 - 7	160	32.32%
7 - 13	44	8.89%
13 - 21	20	4.04%
21 - 35	10	2.02%
35 - 60	4	0.81%
60+	4	0.81%

Table 23, CS3-CS4 Metric 7: Page loading time distribution

Table 23 presents the page loading time distribution, which is almost similar to the CS1-CS2 page loading time distribution (Table 16). The performance of the repository is still considered satisfactory.

Pages / Visit	8.55
---------------	------

Table 24, CS3-CS4 Metric 8: Pages per visit

Average Visit Duration	6:45
------------------------	------

Table 25, CS3-CS4 Metric 9: Average Visit Duration

Finally, Tables Table 24 and Table 25 present some data on the average pages per visit and the average visit duration. This information highlights the fact that the users spent considerable time in the repository and conducted thorough testing.

To conclude, CS3 & CS4 turn out to be an improvement over CS1 & CS2. The platform is proven again to be robust and stable. Features seem to have been improved compared to previous case studies as the Goals achievements show, and there is positive potential for the future. This prospect is also backed by the fact that a very large number of users have tested the platform and provided feedback.

5.4 CS5 & CS6 results

The results from the third instance of the BlogForever platform are fairly positive.

CS5 and CS6 general information is presented in Figure 20. 212 unique visits and 2.075 page views have been performed during the case studies. The average number of pages per visit is 9.79 and the average visit duration is 9:52.

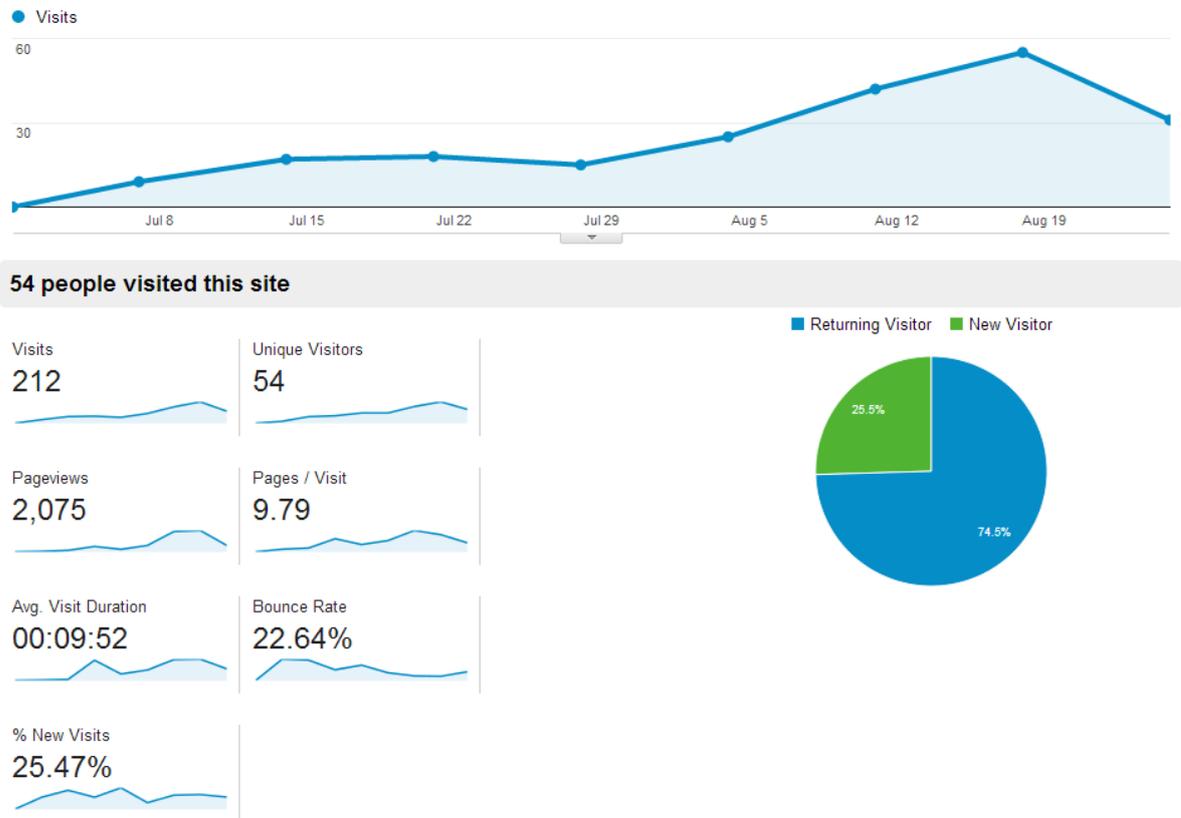


Figure 20, General information on visits for CS5 & CS6

Using Metrics, we are able to isolate specific system variables which are relevant to the aims of the BlogForever Case Studies. The aims of CS5 and CS6 were to highlight border cases, validate the robustness of the platform, check bugs, check the integrated platform and evaluate its robustness. Thus, we are highlighting only relevant metrics on this report.

<i>Number</i>	<i>Percent</i>	<i>Status Code</i>	<i>Explanation</i>
412,620	88.13%	200	OK - The request sent by the client was successful
32,594	6.9%	404	Document Not Found - Requested resource could not be found
12,554	2.6%	302	Moved Temporarily (redirect)
10,223	2.1%	304	Not Modified - Resource has not been modified
75	~0%	500	Internal Server Error
110	~0%	403	Forbidden - Server is refusing to respond to it

Table 26, CS5-CS6 Metric 6: HTTP status distribution

An analysis of Metric 6: HTTP status distribution as presented in Table 30 shows that the number of errors are reduced compared to CS1-CS2 (Table 15) and CS3-CS4 (Table 22). Also, the number of python errors observed in system log files is reduced compared with previous case studies.

<i>Number</i>	<i>Error</i>
13005	traceback.print_stack()
15850	if not self._defer_warnings
15642	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 168, in execute
12565	warn(w[-1], self.Warning, 3)
13534	File "/usr/lib/pymodules/python2.6/MySQLdb/cursors.py", line 82, in _warning_check
13534	File "/usr/local/lib/python2.6/dist-packages/invenio/webinterface_handler_wsgi.py", line 462, in application
13511	rc = cur.execute(sql, param)
13511	result = _check_result(req, obj(req, form))
13511	return root._traverse(req, path, False, guest_p)
13511	return _handler(req)

Table 27, CS5-CS6 Metric 5: Python top code errors

<i>Module</i>	<i>Errors</i>
MySQLdb	35563
invenio errorlib	14112
invenio websearch	44
invenio bibrank	12451
inveniowebsubmit	4700
pyPdf	3501
sets.py	3501
invenio bibdocfile	1609
invenio utils	226
invenio websession	69
invenio webstyle	33

Table 28, CS5-CS6 Metric 5: Errors grouped by python module

Finally, on the performance field, the average page loading time distribution (Table 29) is showing that the performance of the BlogForever platform is not affected when handling the large datasets of CS5 and CS6.

	Page Load Time Bucket (sec)	Page Load Sample	Percentage of total
+	0 - 1	71	8.49% 
+	1 - 3	434	51.91% 
+	3 - 7	253	30.26% 
+	7 - 13	48	5.74% 
+	13 - 21	11	1.32% 
+	21 - 35	8	0.96% 
+	35 - 60	6	0.72% 
	60+	5	0.60% 

Table 29, CS5-CS6 Metric 7: Page loading time distribution

To conclude, CS5 and CS6 system logs reports prove that the BlogForever platform has fewer technical problems than previous case studies. Moreover, it is capable of handling a large number of records without issues.

6 Conclusions

This section summarises the results of the BlogForever Case Studies and those of Work Package 5 in general.

Work Package 5 performed a very complex task, encompassing case studies of different instances and iterations of the BlogForever platform, assessing different implemented features for each case study, analysing different types of blog content of varying complexity and size. Multiple project partners were involved, as well as varied external users who provided the feedback. Overall, the case studies were intended to test the BlogForever platform extensively; generate feedback from users; and minimise system problems.

In our opinion, the chief strength of WP5 resides in defining and working to a clear and common testing methodology, as presented in D5.1 Design of Specific Case Studies [11]. An extensive case study framework was designed, implemented and analysed, creating benchmarks to measure the performance of the BlogForever platform in meaningful ways, with particular focus on the needs of the user. The key stages of this process can be summarised as follows:

1. The research design stage planned an initial set of five general Research Questions (Table 1), to guide the WP5 research team members and keep them focused on the evidence and data needed to validate the BlogForever platform.
2. The research design stage also planned and implemented six case studies of increasing size and complexity.
3. In order to elaborate the Research Questions, a set of ten Themes were defined to help rationalise the outputs of all evaluation reports and connect them with the Research Questions.
4. The case studies data collection relied on the following sources of evidence:
 - a. Previous BlogForever documents and deliverables,
 - b. Internal evaluation as presented in D5.2 Implementation of Case Studies [1],
 - c. External participants interviews and direct observations, as presented in D5.2 and D5.3 User Questionnaires and Reports.
 - d. System monitoring and logs as presented in D5.4 System Logs [10].
5. The final results have been gathered, summarised and analysed in the current report.

Another important point, in our opinion, in the success of WP5 is the interconnection and the synchronisation with WP4 and the project team responsible for the implementation of the BlogForever platform. An iterative development was used for WP4 development processes. According to the DoW: “The design of Task 4.4 will be implemented through iterations. During these iterations, a new modification or add-on will be implemented, tested and documented each time” [12]. In each development iteration, WP5 selected a well-defined set of features to test, evaluate and report back to WP4. This way, both software developers and testers worked together and addressed each others' needs, resulting in continuous improvements to the system and great benefits for the project overall.

WP5 adds further value in describing and recording the process of how the platform is being built and developed. We feel that this work will have value to other similar projects, and to other system developers.

Lastly, WP5 has given voice to the users; it has worked hard to present an understanding of the system from a user's point of view. The internal and external testers have been instrumental in this process, and their opinions and feedback are well represented throughout this study.

At this point, we must highlight an important aspect of the evaluation plan, the use of different versions of the BlogForever platform throughout the case studies. The outcomes presented in WP5 reports represent the testers' opinions and views of the system at the time when each case study was conducted. This means that in many cases, testers have evaluated early or incomplete versions of the BlogForever platform; accordingly, they express views of certain features which were later improved significantly. In this respect, their comments, feedback and scoring do not represent the final state of the system.

6.1 Scoring Results Recap

Tables Table 30 and Table 31 summarise external and internal testing scores obtained through the case studies.

Themes	Score
T1: Using blog records	3.54
T2: System integrity	3.62
T3: Sharing and interaction	3.70
T4: Searching	3.46
T5: Access	3.72
T6: Data integrity	3.56
T7: Preservation	3.78
T8: Functionality	3.63
T9: System navigation	3.60
T10: System terminology	3.51
Average	3.61

Table 30, External Testing Scores Summary

Themes	Score
T1: Using blog records	3.57
T2: System integrity	4.00
T3: Sharing and interaction	3.09
T4: Searching	3.83
T5: Access	3.40
T6: Data integrity	3.75
T7: Preservation	3.00
T8: Functionality	4.11
T9: System navigation	3.00
T10: System terminology	3.00
Average	3.66

Table 31, Internal Testing Scores Summary

Some interesting observations can be summarised as follows:

- The average score in internal and external testing is almost equal, 3.66 and 3.61. This score is rather favourable as 3 equates to "Most areas worked as expected" and 4 equates to "All work as expected".
- On average, the scores of all external testing scores are very consistent, with a minimum of 3.46 in T4: Searching and a maximum of 3.78 in T7: Preservation

(range 1 to 5). On the other hand, internal testing scores vary more, with a minimum of 3 in T9: System navigation, T10: System terminology and T7: Preservation. The maximum value is 4.11 for T8: Functionality.

- There seems to be consensus on the scoring of most themes. The difference between the external and internal testing is less than 1 for all Themes, strengthening the outcomes of the evaluation.

6.2 Research Questions Outcomes

Regarding the original Research Questions, we can summarise our findings as follows:

RQ1: What are the particular **problems** the implementation is facing? Or are the BlogForever software implementation processes an **overall success**?

The answer to RQ1 lies in the general outcomes of the evaluation as it is trying to evaluate the overall success of the BlogForever software implementation. Nevertheless, we tried to focus on some specific aspects of the evaluation results, which we believe apply directly to RQ1:

- Looking into the general information of the case studies as presented through the system logs (Table 12, Table 19), we see that the number of visitors and page views is substantial. These statistics demonstrate the rigour of our testing process: multiple tests were made by many users. It also shows the platform is capable of handling a large number of users and requests.
- In addition, evaluating the system logs metrics and especially Metric 5: Python error codes and Metric 6: HTTP status distributions, we see that very few system errors have occurred considering the testing process.
- Finally, Theme 8: Functionality and Theme 2: System integrity scores are above the average in internal testing results and near the average in external testing results.

RQ2: Are complex BlogForever platform search strategies **working efficiently** when high levels of content are available within the BlogForever platform?

The last two case studies are characterised by the large volume and complexity of the blogs in scope. The system logs evaluation of CS5 and CS6 (Section 5.4) show that the BlogForever platform is working efficiently.

RQ3: How **useful** is the BlogForever platform as a whole?

RQ3 is aligned with the following Themes: T2: System integrity, T6: Data integrity, T7: Preservation and T8: Functionality. Their internal and external testing scores are presented in Table 32.

Themes	External Testing Score	Internal Testing Score
T2: System integrity	3.62	4.00
T6: Data integrity	3.56	3.75
T7: Preservation	3.78	3.00
T8: Functionality	3.63	4.11

Table 32, RQ3 related Theme scores

Furthermore, the following observations are also relevant to RQ3:

- The System Logs Metrics relevant to T2: System integrity are M5: Number of python errors, M6: HTTP status distribution, M7: Page loading time distribution [10]. As we already presented in RQ1 discussion, these Metrics are very favourable as they present a system of high integrity and few errors. Also, one of the highest scores in any Theme and context goes to T2: System integrity (4.00).
- As we have already presented in the conclusions of D5.2 regarding preservation [1]:
 - The evaluation demonstrates that all the data captured by the spider is being ingested into the system; however, improvement is necessary for capturing additional contextual metadata about the crawl and the blog.
 - More description is needed with regard to the content of the object (e.g. topic, language, etc.) and to technical aspects (e.g. formats). (This has improved since CS1).
 - Content retrieved via APIs is stored in two different databases as part of the preservation strategy.
- T8: Functionality is considered very positively by the majority of users. The maximum score of all Themes in internal and external evaluation is T8: Functionality (4.11). In addition, there are numerous positive user comments in D5.3 User Questionnaires and Reports, praising the system's functionality.

RQ4: Does the use of the BlogForever repository lead to **successful** results for the different users?

RQ4 is aligned with T4: Searching and T5: Access. Their internal and external testing scores are presented in Table 33.

Themes	External Testing Score	Internal Testing Score
T4: Searching	3.46	3.83
T5: Access	3.72	3.40

Table 33, RQ4 related Theme scores

Furthermore, the following observations are also relevant to RQ4:

- Regarding **T4: Search**, although it has powerful features, the GUI of the search function is neither intuitive to use nor consistent, as many users of the External Evaluation have noted in various ways such as:
 - *"Options under the search input are incomprehensible."*
 - *"in a phrase search with 2 words done that there were no results, there was not a clear message that there were no results and the suggestion of results with each of the words confused the user because there was no explanation text."*
 - *"blog post search results should have the same format with the blog search results (blog results shows the available files (.jpeg, .txt, xml_mets,) but post results do not contain them)."*
 - *Search results: records should be hyperlinked (users clicked randomly either on "HTML" or "detailed record")*
 - *Some search fields not relevant*
 - *Advanced Search form is overloaded with fields.*
- The outcomes of **T5: Access** evaluation are mixed in the conclusions of D5.2 [1]:
 - Some issues related to access to the blog records and representation of dissemination copies of the content were identified.
 - The system captures the layout and overall look of blogs as expected.

- The bookmarking, export function and the use of UTF to enable multilingual content operate as expected. However, the translation features perform inconsistently.
- Presented content is harmonised in the repository and is consistently displayed across many different browser types.
- Extraction of content into MARC XML and DC XML performs as expected, but recommendations for including METS/METS XML for import and export are made.
- Extraction as PDF or Image does not perform as expected.
- Support for OpenURL does not perform as expected.
- Navigation of blogs by topics does not perform as expected.
- Improvements are needed for the readability of the statistics graphs.

RQ5: How **user friendly** are the BlogForever platform functions for the different designated blog communities?

RQ5 is aligned with the following Themes: T1: Using blog records, T3: Sharing and interaction T9: System navigation and T10: System terminology. Their internal and external testing scores are presented in Table 34.

Themes	External Testing Score	Internal Testing Score
T1: Using blog records	3.54	3.57
T3: Sharing and interaction	3.70	3.09
T9: System navigation	3.60	3.00
T10: System terminology	3.51	3.00

Table 34, RQ5 related Theme scores

Furthermore, the following observations are also relevant to RQ5:

- **Regarding T1: Using blog records:** The internal evaluation in D5.2 demonstrated that most of the aspects related to the usage of blog records operate as expected. [1]. The external evaluation score is just around average.
- **Regarding T3: Sharing and interaction:** The internal evaluation showed that not all features performed as expected while external testing score is high (3.70).
- **Regarding T9: System navigation:** The external testing score is average (3.60) but there are numerous suggestions about it in users' feedback:
 - *"Bigger tabs and bigger font-size."*
 - One tester noted *'People who don't use technology can get quite scared and don't need much of an excuse to say bye bye.'*
 - *"Tabs and breadcrumb trail font is far too small. Think about people with visual disabilities. Add 'your notifications' to tab to avoid email glut of notifications"*.
 - Dashboard features lauded but lack of instruction of functionality e.g. *"what happens when you shut down one of the boxes on dashboard? How can you bring it back?..figured it out but guidance needed."* Finding the dashboard is not intuitive, most found it a bit confusing.
- **Regarding T10: System terminology:** There are a lot of issues and suggestions about this Theme which lead us to believe that it could be improved considerably:
 - Terminology is unclear to the layperson, i.e. a non developer e.g. *"What does 'similar records' mean? Basic understanding/explanation of terms and how to search and what they will find on searching."*

- All testers reported that finding where to **register** was unclear and most spent 1 minute locating it at 'login'. Suggest 'login/register' title to this function. Keep in same place.
- Droplist of options from Personalize tab, it is clear that some are not relevant, e.g. "' loans', groups? Please explain or have some help section to expand."

To conclude, we consider the rating of all Themes to average out between 3 ("Most areas worked as expected") and 4 ("All work as expected"). Any deviations between the different themes, evaluation methods and case studies are not significant. From these scores, we can conclude that the majority of users are satisfied with the performance of the BlogForever platform. The trend of our success in WP5 indicates that any outstanding issues embodied in the Research Questions will be addressed by future project work.

In general, we may with good confidence assert that the BlogForever Case Studies show the credibility of the BlogForever approach for a blog aggregation, preservation, management and dissemination platform.

7 References

1. **Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent, George Gkotsis, Karen Stepanyan, Morten Rynning, Stella Kopidaki.** *BlogForever D5.2 Case Studies Reports*. s.l. : BlogForever, 2013.
2. **S. Arango-Docio, P. Sleeman, H. Kalb.** *BlogForever D2.1 Survey Implementation Report*. 2011.
3. **K. Stepanyan, M. Joy, A. Cristea, Y. Kim, E. Pinsent, S. Kopidaki.** *BlogForever D2.2 Weblog Data Model*. 2011.
4. **H. Kalb, Y. Kim, P. Lazaridou.** *BlogForever D2.3 Weblog Ontologies*. 2012.
5. **K. Stepanyan, G. Gkotsis, E. Pincet, V. Banos, R. Davis.** *BlogForever D2.6 Data Extraction Methodology*. 2012.
6. **Yunhyong Kim, Seamus Ross, Karen Stepanyan, Ed Pinsent, Patricia Sleeman, Silvia Arango-Docio, Vangelis Banos, Ilias Trochidis, Jaime Garcia Llopis, Hendrik Kalb.** *BlogForever D3.1 Preservation Strategy*. 2012.
7. **Hendrik Kalb, Paraskevi Lazaridou, Matthias Trier.** *BlogForever D3.2 Interoperability Prospects*. 2012.
8. **Hendrik Kalb, Nikolaos Kasioumis, Jaime García Llopis, Senan Postaci, Silvia Arango-Docio.** *BlogForever D4.1 User Requirements and Platform Specifications*. 2011.
9. **Hendrik Kalb, Paraskevi Lazaridou, Matthias Trier.** *BlogForever D4.2 Weblog Spider Component Design*. 2012.
10. **Rynning, Morten.** *BlogForever D4.3 Initial Weblog Spider Prototype*. 2012.
11. **Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent, Tracie Farrell.** *BlogForever D5.1 Design and Specification of Case Studies*. s.l. : BlogForever, 2011. Deliverable.
12. **Rynning, Morten.** *BlogForever D4.6 Final Weblog Spider Component*. 2013.
13. **Senan Postaci, Alper Cinar, Gokce Banu Laleci, Jaime Garcia Llopis, Raquel Jimenez Encinar, Vangelis Banos, Apostolos Papadopoulos.** *BlogForever D4.5 Initial Weblog Digital Repository Prototype*. 2013.
14. **J. Garcia Llopis, R. Jimenez Encinar, S. Postac, A. Cinar, V. Banos, H. Kalb.** *BlogForever D4.7 Final Weblog Digital Repository Component*. 2013.
15. **Vangelis Banos, Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent, George Gkotsis, Karen Stepanyan, Morten Rynning, Stella Kopidaki.** *BlogForever D5.3 User Questionnaires and Reports*. 2013.
16. **Banos, Vangelis.** *BlogForever D5.4 System Logs*. 2013.
17. **BlogForever.** *Annex I - Description of Work*.
18. **Vangelis Banos, Silvia Arango-Docio, Patricia Sleeman, Ed Pinsent.** *BlogForever D5.5 Case Studies Comparative Analysis and Conclusions*. 2013.