

An Endless Tail

**The National Library of Norway's "High North" pilot project –
Evaluation as at 1 October 2008**

Contents

0	Summary	p. 3
1	Background to the pilot project	p. 4
2	Description of functionality. Definitions	p. 4
3	Experiences	
	3.1. Description of statistical tools	p. 6
	3.2. Key figures	p. 6
	3.3. Usage by genre and subject	p. 8
	3.4. Usage by year of publication	p. 9
	Relative figures – number of works:	
	3.5. number of page imprints	p. 10
	3.6. The ten most active users	p. 10
	3.7. The ten most popular titles	p. 13
	3.8. Use of the service : library loans	p. 13
	3.9. Use of the service : sales	p. 15
	3.10. Interview survey	p. 16
4	Overall assessment	p. 18
5	The road ahead	p. 20

Project steering committee

National Librarian Vigdis Moe Skarstein (National Library of Norway), Secretary-general Trond Andreassen (Norwegian Non-fiction Writers and Translators Association), Secretary-general Mette Møller (Norwegian Authors' Union), Advocate Mathias Lilleengen (Norwegian Publishers Association)

Project working committee

Project manager Stig Bang (National Library of Norway), General manager Thomas Nygaard (PT), General manager Mads Liland (Licensing and Collecting Society LINO)

Special consultant

Professor Ragnar Waldahl (UiO)

"... older content tends to score higher because it's had longer to accumulate incoming links. In other words, search inverts the usual priority of content; older is often the better."

[Chris Anderson – "The Long Tail", 2006]

0. Summary

The National Library of Norway and representatives for Norwegian rights holders have collaborated on a pilot project in which copyrighted Norwegian works are made available in a digital format via the internet. The High North project, which was launched on 24 April 2007, will run until the end of 2009. In addition to disseminating knowledge about the High North, the purpose of the project has been to gain genuine experience of the use of digitalised books and journals online. This report is based on the statistical analysis of the behaviour of users of the online service, of library lending figures and of sales statistics during the same period, in addition to findings from an online questionnaire. Below are some of the key findings of the analysis:

The whole tail wags

The study shows there was a demand for all subject areas and genres, and a full 87 per cent of titles were accessed.

Positive reading experience

Fifty-five per cent of people asked rated their experience of reading books and journals online from "OK" to "Very good".

Time generates access

From April 2008 until October 2008, the total number of pages displayed grew by 59 per cent, despite the fact that the service itself had experienced minimal growth, and even though no special initiatives had been implemented to raise awareness of the service.

The service is used as a library, not as a reading room

On average each user views five unique pages in each book and two pages in each journal, often visiting the service several times and viewing the same pages, bringing the average number of pages displayed to fourteen for books and five for journals.

Everything is growing – some things are growing more than others

Usage of all the titles included in the service has increased during the report period, but, relatively speaking, there was greatest growth among the most popular titles.

Search engines

When Google and other search engines were given access to the service's metadata in January 2008, the number of objects accessed each month increased by 61 per cent compared with figures for 2007.

Older material is not dead

Sixty-five per cent of the titles accessed are 11 years or older relative to the year of publication.

1. Background to the pilot project

Norwegian writers, publishing firms, journal publishers, other copyright holders and the National Library of Norway all have a common goal, which is to enable access to literature and other copyright protected material, and to give the public an insight into and knowledge of the diversity in the Norwegian cultural heritage.

As part of this goal, the National Library of Norway and representatives for Norwegian copyright holders¹ have for the last two years collaborated on a pilot project in which Norwegian works are made available in a digital format via the Internet within a more closely defined subject area – **the High North**. The aim of this pilot project is, among other things, to gain experience with technological solutions, problems with links, agreement mechanisms, usage patterns etc., which can form the basis for any future agreements on the digital dissemination of copyright protected material.

The High North was chosen as a basis for the pilot project due to current interest in this area. The Norwegian government's focus on the High North as a geopolitical area opens for many problem issues and for material from many sources. Collecting the most relevant material in one location would represent an important contribution to knowledge about the High North, particularly since the works could be presented in full-text versions on the internet. The sample consisted of works of fiction and non-fiction evenly distributed according to year of publication, with the most recent titles being publications from the previous year.

The National Library of Norway was responsible for selection of the sample, and for digitalisation and technical solutions, while the licensing organisations handled the clearing of rights to the copyrighted material. The comprehensive collection of links was produced with the assistance of the Barents Secretariat and the Fellowship Programme for Studies in the High North at the University College of Bodø.

The terms and conditions governing collaboration between the National Library of Norway and the licensing organisations have been set forth in a separate agreement (see <http://www.nb.no/nordomradene/>).

The pilot project was launched on 24 April 2007. One year later experiences of the project were summarised in a quantitative analysis of user patterns as at 24 April 2008 (see <http://www.nb.no/nordomradene/Nordomradene-midtveisrapport.pdf>). In *this* evaluation, the "midway report" has been expanded on and all experiential documentation has been elaborated on and updated as at 1 October 2008.

2. Description of functionality. Definition

Definitions

In this project report, the following terms are defined in the following manner:

User means a person with a unique IP address.² As regards the term **number of unique IP addresses** (e.g. table 1), adjustments have been made for users who have accessed both books and journals.

¹ Copyright holders are represented by the Norwegian Publishers Association, the Norwegian Authors Union, the Norwegian Non-fiction Writers and Translators Association, the Norwegian Critics Association and the Licensing and Collecting Society LINO. In a separate agreement with the Norwegian Visual Artists Copyright Society, the Norwegian Organisation for Visual Communication, the Society of Fine Art Photographers in Norway and the Norwegian Photographers' Union, LINO has the authority to enter into an agreement on behalf of the members of these organisations in the pilot.

² For a definition of an IP address, see section 3.1 below.

A **visit** occurs when a user opens a digital document in order to view it. If a user first "visits" a book and then switches to a journal, this is registered as two visits.

A **viewing** of one or more pages of a book or journal occurs when a user "visits" a work and studies the page(s) on a computer terminal. A page is "viewed".

An **object** is a collective term for all material offered by the service (books, journals, photographs, radio programmes). In this report, the term object is used in section 3.6 and sporadically in section 3.2. The analyses otherwise relate to books and journals.

Number of unique pages means the total number of pages the user has at his or her disposition.

Number of unique pages viewed means the number of unique pages users have *actually* used.

Number of pages viewed as a separate category shows the fact that a unique page has been viewed several times, by one or more users.

Journal means an article in a journal or chapter in a book where there is more than one contributor (collective work).

This report may contain some minor inconsistencies (e.g. in some cases metadata and statistical data may be incongruent), but this has no particular impact on the results. In the National Library's catalogue, journals are registered at title level (*Ottar*, *Geo*, *Fauna* etc.), while the title of the article is not registered. In several cases, one edition of a journal may contain more than one article for which the rights have been cleared, which means that the number of articles is actually slightly higher than the number of "journals" that has been logged and analysed.

Searching and retrieval of High North material

There are two ways of accessing the material, either by the portal <http://nb.no/nordomradene> or by searching the National Library of Norway's collection. This is an integrated search that trawls many different databases of source material, irrespective of material type and subject area. All material that comes within the realm of the High North project is marked as such in the National Library's catalogue, The search engine recognises this and establishes a separate "High North collection" online. The High North collection consists of 395 books, 248 journals, 401 web pages, 1547 photographs and 19 radio programmes.³ A search can be limited to only include part of the index.

Full-text display of books and journals

Books in NBdigital are scanned using optical character recognition (OCR) and structurally analysed. Supplementary metadata is also added. Any text retrieved using OCR is indexed in the National Library's search engine together with the metadata. The High North material is not treated any differently.

It is thus possible to search in the traditional manner for title and author (metadata), and at the same time search for words and phrases within the content itself (free-text search). If a search word is found in the text, the user is taken to the pages in the book where the text was found and he or she can then browse the work.

An automated structural analysis is also carried out in which any list of contents is annotated, and page numbers in the book are verified to ensure that the digital rendering retains the original pagination.

Users may leaf through works, jump to a particular page number and display the work at different resolutions and in different ways. The pages shown are displayed as

³ The High North portal <http://nb.no/nordomradene> contains links to more than 200 websites/online documents with a content comprising core material about the High North. Many of these works were born digitally and are found in open research archives outside the National Library of Norway. For this reason, they are not included in the index of digitised High North projects, cannot be searched in NBdigital, and will thus have not influence on the figures in this study.

digital images (JPEG format) of genuine pages of books, presented as individual images. No provision has been made for the printing of pages.

Journals are scanned in their entirety, but access is limited to ensure that users may only read articles where the rights have been cleared.

Bibsys users can go directly from viewing a book to ordering it through Bibsys. Users who wish to buy the book can click on a link which takes them to the sales service operated by the licensing organisations and buy the book or order a print-on-demand version of works that are not otherwise available.

The High North portal

When a user search produces hits in material marked "High North" in the catalogue, or the search word is found in a pre-defined list of High North-related subject words, a "super hit" is activated that gives access to the High North portal (<http://www.nb.no/nordomradene/>), where the High North concept is placed in a broader perspective, along with relevant news updates, articles, presentations and a large collection of links to relevant High North resources throughout the world.

Visibility

Digital objects in the National Library of Norway are not open for indexing by the major search engines. For books and journals in the High North collection, open web pages were established to make the metadata available to Google and other search engines. The material has therefore been searchable outside NBdigital after 16 January 2008.

With the exception of editorial news items in the press, the ether media and specialist journals, the service has not been advertised in newspaper adverts, television advertising or other forms of paid marketing.

3. Experiences

3.1 Description of statistical tools

Access to information covered by the agreement is based on the National Library of Norway's ordinary digital library, NBdigital. The service provided is a relatively standard web-based service, providing access to most types of information in digital form available from the National Library's collection.⁴

All access to digital objects is logged. In this connection, a digital object comprises, for example, one audio clip, one page of a book or one photograph. The logs contain information such as the user's IP address,⁵ time of access and a reference to the object viewed.

Logs are transferred on a daily basis from a flat-file database to a relational database that has been specially developed in order to provide statistical data. The database only retains log information and thereby provides effective, flexible and dynamic access to log data. Users are rendered anonymous as their IP addresses are changed to untraceable identity factors. Nevertheless, each user remains unique, allowing one to relate user behaviour to each individual anonymous IP address.

It should be stressed that one individual IP address registered in a log may represent more than one user since IP addresses are allocated dynamically or because it may represent a network of private addresses behind one IP address (NAT).⁶ For this reason, there is some uncertainty as to whether one IP address represents one user.

On the basis of this data, extrapolations, comparisons and some calculations have been made. All functionality has been tailored for the specific purpose and developed by the National Library of Norway. Much of what is presented is based on limitations made

⁴ As at 1 October 2008 the digital collection consists of 200,000 newspapers, 365,000 images, 73,000 books, 500 films, 1,000 hours of music, 300,000 hours of radio programmes and 250 journals.

⁵ IP address (Internet Protocol address) is the address used on the internet for units that may be addresses, based on a combination of numbers enabling one to navigate the internet.

⁶ NAT, Network Address Translation, is used to change IP addresses in routers and firewalls. Using this method, many unofficial/private IP addresses can share one official IP address.

by the user, and processing must occur at the time of access. The basis for analysis will never be older than approximately one day.

3.2 Key figures

Description of the analysis

The point of this analysis is to describe the overriding figures and to chart any changes from the review made in April, and to update a number of percentages and changes in relation to the figures available as at 1 October.

Results

The figures in table 1 show that 87 per cent of the total number of titles available in the service have been accessed. This represents a 12 per cent increase in the period from April to October 2008. The figures for books and journals are 89 per cent and 85 per cent, respectively. The share of unique pages viewed comprises 19 per cent of the number of unique pages available. The number of pages overall was 59 per cent higher in October than in April. For books the increase is 66 per cent; for journals this figure is 26 per cent.

In January 2008, the service's metadata was opened to, and indexed by, search engines. After this time, the number of *objects* accessed per month in 2008 was found to be 61 per cent higher than the average monthly figures for 2007. (This figure is obtained by analysing the basic data, but has not been included in the table below.)

Key figures Category	24.04.08		01.10.08		Change
	Values	%	Values	%	
Number of titles	624		643		3 %
* of which books	394		395		
* of which journals	230		248		
Number of unique pages	81473		82341		1 %
* of which books	71425		71748		
* of which journals	10048		10593		
Number of unique IP addresses	2482		3318		34%
* of which books	2013		2683		
* of which journals	801		878		
* share of books + journals	13.40 %		7.32 %		
Number of visits	4605		5280		15 %
* of which books	3524		4150		
* of which journals	1081		1130		
Number of titles accessed	500	80 %	561	87 %	12 %
* of which books	308	78 %	350	89 %	
* of which journals	192	83 %	211	85 %	
Number of unique pages viewed	12157	15 %	15684	19 %	24 %
* of which books	10535	15 %	13955	20 %	
* of which journals	1622	16 %	1729	16 %	
Number of pages viewed	27753		44055		59 %
* of which books	23014		38099		66%
* of which journals	4739		5956		26%
Relative figures (average)					
Number of pages viewed : number of visits					
* books	7.10		9.18		
* journals	4.38		5.27		
Number of pages viewed : number of unique IP addresses					

* books	12.43	14.20
* journals	5.92	5.27
Number of unique pages viewed: number of unique IP addresses		
* books	5.23	5.20
* journals	2.02	1.97
Number of unique pages viewed : number of visits		
* books	2.99	3.36
* journals	1.50	1.53

Table 1: Key figures

Evaluation

The figures for October reinforce the trend from April since the lion's share of the titles (87%) offered by the service have been accessed, representing an increase of 12 per cent. In the opinion of the project management, these figures confirm that both the service and its content are considered by the general public to be attractive (more about this later). The fact that the number of pages viewed has increased by 59 per cent provides further confirmation of this, in spite of the fact that the growth in users of the service has been minimal during the same period. Of, if you wish: the service's user frequency increases with the length of the project period. That the number of unique pages viewed is barely 20 per cent conceals great variations, as we shall see below. Finally, one may note that making metadata available to the search engines has contributed significantly to the increase in user frequency.

3.3 Usage by genre and subject

Description of the analysis

In April, the 100 most visited titles were analysed – books and journals combined. In the October analysis, the April figures have been broken down by genre, although still based on the 100 most visited titles. The October figures are based on all titles available from the service.

Results

As one can see (table 2), there has been a certain shift in the figures from April to October, as evidenced by the "increase" in the share of journals, and a certain amount of levelling off between subject areas. For example, the dominant position of history in April has shifted in favour of the social sciences, travel literature and natural sciences.

Genre and subjects		
Category	24.04.08	01.10.08
	*	
Journals	31 %	38 %
# of which fiction literature	0 %	0
# of which non-fiction literature	100 %	100 %
Subject		
(Cultural) history, biographies	65 %	56 %
Social sciences, strategic studies	26 %	8 %
Nature, outdoors, travel	6 %	20 %
Language, literature	3 %	16 %
Books	69 %	62 %
# of which fiction literature	13 %	17 %
# of which non-fiction literature	87 %	83 %
Subjects		
(Cultural) history, biographies	70 %	52 %

<i>Social sciences, strategic studies</i>	13 %	20 %
<i>Nature, outdoors, travel</i>	7 %	24 %
<i>Language, literature</i>	10 %	4 %
Books and journals together	100 %	100 %
# of which fiction literature	9 %	10 %
# of which non-fiction literature	91 %	90 %
Subjects		
<i>(Cultural) history, biographies</i>	68 %	54 %
<i>Social sciences, strategic studies</i>	7 %	15 %
<i>Nature, outdoors, travel</i>	17 %	22 %
<i>Language, literature</i>	8 %	9 %
* The figures for April are based on the 100 most accessed titles (ordered by IP addresses).		

Table 2: Usage by genre and subject.

Evaluation

Based on the figures available, it would, in the opinion of the project management, be unwise to draw any unequivocal conclusions regarding the causes of the trends documented. Both the time factor and the very different distributions of subject areas and genres on the "popularity" scale may be possible explanations.

3.4 Usage by year of publication

Description of the analysis

In April, 280 titles were analysed – for books and journals combined. These figures for April have now been broken down by genre, though still based on the same 280 titles. In the figures for October, all the titles in the service have been included – at genre level and as a whole.

Usage by year of publication		
Category	24.04.08 **	01.10.08
Journals		
Year of publication *	0 %	0 %
1 - 2 years	9 %	6 %
3 - 6 years	8 %	11 %
7 - 11 years	18 %	16 %
> 11 years	65 %	67 %
Books		
Year of publication *	0 %	0 %
1 - 2 years	8 %	7 %
3 - 6 years	11 %	15 %
7 - 11 years	16 %	14 %
> 11 years	65 %	64 %
Books and journals combined		
Year of publication *	0 %	0 %
1 - 2 years	8 %	6 %
3 - 6 years	12 %	14 %
7 - 11 years	17 %	15 %
> 11 years	62 %	65 %
* Defined as 2007. No works included in the service		
** The figures for April are based on the 280 most accessed titles (arranged by IP addresses).		

Table 3: Usage by year of publication.

Results

In October, 65 per cent of titles available through the service were 11 years' old or more, which is slightly higher than in April. It may otherwise be noted that there are only small differences in this analysis between books and journals.

Evaluation

That the figures for journals and books are largely concurrent may be regarded as surprising. A U.S. study⁷ to which the project management has had access would tend to indicate a stronger actuality profile in respect of the digital use of journals compared with the similar use of books. The reason for this difference may be that the cited study analysed use of digital works among students; the actual study situation, professional orientation and age composition of respondents would differ here from users of the High North service. On the other hand, the figures on which this present evaluation are based appear to indicate that the service is used for a number of purposes, and is perceived as a diverse and "open" resource, in the same way as a library functions and is supposed function. The "footprint analysis" below (see section 3.6) would appear to support this assumption.

3.5 Relative figures – number of works : number of pages viewed

Description of the analysis

The aim of this study is to examine how large a share of the number of pages viewed can be attributed to the 2, 20 and 100 most accessed titles. In April, books and journals were analysed in combination; in table 4, these figures have been broken down by genre.

Number of works: number of pages viewed		
	24.04.0	01.10.08
Category	8	8
Journals		
2 works	8 %	8 %
20 works	33 %	27 %
100 works	76 %	77 %
Books		
2 works	7 %	7 %
20 works	30 %	30 %
100 works	66 %	75 %
Books and journals combined		
2 works	5 %	1 %
20 works	25 %	27 %
100 works	60 %	64 %

Table 4: Number of works : number of pages viewed based on the 100 most viewed titles (ordered by IP address).

Results

⁷ "There was a focus of interest in the most recent articles. Twenty-two per cent of pages viewed were published in the current year, 42 per cent were one to three years old, 24 per cent were four to seven years old and 12 per cent were over seven years old." Quoted from David Nichols et al.: "What does usage data tell us about users?" in the programme prepared for the Online Information Conference, London 4-6 December 2007, p. 84.

Taken as a whole, the October figures show that the shares for the three groups are 1 per cent (slightly lower than in April), 27 per cent and 64 per cent (slightly higher than in April). If one looks at the subgroups, it can be seen that the 100 most accessed titles comprise 77 per cent (journals) and 75 per cent (books), respectively.

Assessment

It may seem surprising that the 100 most visited titles (books) comprise 75 per cent of the total number of pages viewed, and that this figure grew by almost 10 per cent in the period from April to October 2008. This must mean that the other book titles accessed have received correspondingly less interest. In the opinion of the project management, this is because we face a kind of "normal distribution", where the 100 most visited titles "drag their tails behind them"; this may be a natural conclusion since the greatest growth in traffic is seen among the 100 most accessed titles. The relative increase in frequency of the remaining titles is still there, but the effect is a shift in the percentage shares. This phenomenon is supported by the fact that the number of pages viewed during the period April to October increased by 59 per cent (see table 1 above). It would thus appear that the viewing of titles along the entire long tail increases, and that those titles that were accessed the most grew the most.

3.6 The ten most active users

Description of the analysis

In table 5, the ten most active users have been isolated and the number of objects they have accessed has been charted, along with the number of unique pages they have accessed and how many pages have been viewed in total. Note that this analysis also includes these users' interest in photography and radio broadcasts ("Other"), which also form part of the service. Finally, an analysis has been made of the figures in the table in order to "follow" each user through the material.

Results

It can be seen from the figures that the ten users have extremely varied usage patterns; while some access a large number of titles, others are interested in only a small handful of titles. It can also be seen that interest in objects other than books and journals is limited in this user group.

The ten most active users					
User	No. of unique pages	No. of pages viewed	Viewed :	No. of objects accessed	Percentage of objects accessed
User 6932	799	1278	2.62 %	31	1.30 %
Book	640	1049			
Journal	149	218			
Other	10	11			
User 11142	763	763	1.56 %	383	17.34 %
Book	414	414			
Journal	349	349			
Other	0	0			
User 9059	755	755	1.50 %	378	17.11 %
Book	406	406			
Journal	349	349			
Other	0	0			
User 10758	340	588	1.20 %	19	0.86 %
Book	324	565			

	Journal	16	23			
	Other	0	0			
User 7173		396	534	1.09 %	3	0.14
	Book	396	534			
	Journal	0	0			
	Other	0	0			
User 4727		465	486	0.99 %	65	2.94
	Book	407	424			
	Journal	46	50			
	Other	12	12			
User 3		299	505	1.03 %	76	3.44 %
	Book	228	416			
	Journal	60	76			
	Other	11	13			
User 3974		245	468	0.96 %	10	0.45 %
	Book	244	466			
	Journal	0	0			
	Other	1	2			
User 9460		425	449	0.92 %	265	12.00 %
	Book	222	232			
	Journal	203	215			
	Other	0	0			
User 19727		391	427	0.87 %	264	11.95 %
	Book	390	241			
	Journal	219	186			
	Other	171	0			
TOTAL		4878	5573	11.41 %		
	Book	3671	4747			
	Journal	1391	1521			
	Other	205	38			
In the service		16976	48862	100 %	1124	100 %
	Book	13955	38099	77.97 %	357	31.73 %
	Journal	1729	5956	12.19 %	214	19.04 %
	Other	1292	4807	9.84 %	553	49.20 %

Table 5: The ten most active users

The High North website is built up so as to give users many ways of navigating the material available. This is also reflected in the actual use of material.

The project group has studied the ten most active movements through a random selection of titles that each individual has visited. They access a copyrighted work either by clicking on a link on the actual title of the work or by clicking on a link to the page of the work where their search engine returned a result. Regardless of which of these two ways used to arrive at the specific digital object, the user's subsequent movement within the work itself appears to be more dependent on the intentions of the user in accessing the work than the service's functionality.

The length of time between a user accessing one object and the next indicates whether he or she is only browsing or is actually reading the text available on each page accessed. Some users quickly leaf several pages back and forth in a work before moving on to another work and/or leaving the service, while other users work their way more systematically backwards or forwards through the work selected. Among those users who

have had a more systematic approach to the works, we also see that both lists of contents and reference lists are frequently accessed.

Among the ten most active users, only one user has been found who on one occasion has most likely read one work from cover to cover (Ailo Gaup's novel *Natten mellom dagene [Night between days]*, Gyldendal 1992). In most cases, users browse through parts of a work, either systematically or more randomly. Among the more systematic users, we find instances where it is likely they have found in one work a reference to another work, and have then "jumped" to the work referred to before returning to the original work or moving on to another work. We also find that users who have accessed journals are more likely to read/browse articles in their entirety than users who access books.

The website contains many different forms of objects in addition to books – journals, photographs, audio recordings etc. What is striking is that users primarily visit texts – either pages of journals or book pages.⁸

Evaluation

The background data for this table shows that the fewer works a user visits, the more pages he or she reads. One example: user no. 7173 visited three titles; one of these is a novel that has probably been read from cover to cover. The amount of time spent on this activity, which can be documented, substantiates the assumption that the novel was read in its entirety.

In so far as one can draw any general conclusions from an analysis of user behaviour, it must be that users have used the service in the same way they would use book collections in libraries. They select books and journals that appear interesting and browse through them in order to decide whether they suit the user's wishes and needs. The High North website is a web-based library. Only to a limited extent is the website used as a "reading room" – exceedingly few users read large tracts of text on screen.

3.7 The ten most popular titles

Description of the analysis

The analysis presented in table 6 describes the use of the ten most accessed titles as at October 2008, and analyses how users navigate these titles.

Results

In April, all ten titles were works of non-fiction, while in October two works of fiction were included, in second and third place. And in both cases, each unique page had been accessed, with the two works being accessed by 159 and 133 unique IP addresses, respectively. Historical works dominated the list in April, with the addition of two works from the subject group language and literature.

The ten most accessed titles						
Title	Unique pages	Unique pages viewed	%	No. of pages viewed	No. of unique IP addresses	No. of pages viewed/IP
I nasjonalstatens tid	583	357	61 %	1439	222	6.5
Pan	101	102	100 %	1065	159	6.7
Benoni	165	163	100 %	1298	133	9.8
Norges religionshistorie	587	168	29 %	948	122	7.8
I kongenes tid	477	326	68 %	790	90	8.8
Historien om Nord-Norge	517	198	38 %	581	88	6.6
No. polarhist. Rikdommen	605	112	19 %	311	86	3.5

⁸ One user has clearly been interested in illustrations/photographsm but has searched for these in book illustrations or reproductions of photographs in books.

Knut Hamsun	29	29	100 %	368	83	4.4
Mot rikare mål å trå	383	127	33 %	1254	80	15.7
Norge	457	93	20 %	238	77	3.1
TOTAL	3904	1675	43 %	8288		
In the service	82341	15684	20 %	44055	3318	13.3
Percentage	4.70 %	11 %		19 %		

Table 6: The ten most accessed titles

Evaluation

As in the analysis presented in the previous section (section 3.6 above), this analysis also documents considerable variation in usage patterns. Based on the footprint analysis, it may generally be said that works of fiction are leafed through and read more than is the case with works of non-fiction (the assessment made in the previous section – one searches for sources of knowledge that fit the users' needs – this applies in particular to works of non-fiction). Finally, the footprint analysis documents that Knut Hamsun's *Pan* was read in its entirety.⁹

3.8 Use of the service : library lending

Description of the analysis

The project group has obtained overviews of the lending of literature in the High North project from Bibliofil (public libraries) and Bibsys (specialist libraries). These overviews have been matched and compared.¹⁰ This analysis concerns only books, since it is not possible to extract lending data for individual articles in journals.

Results

The overviews show an increase of approximately 24 per cent in the lending of High North literature from libraries from the start of the project up until 1 October 2008. Further, we see that there has been an increase in lending for 58 per cent of all 391 titles, while approximately 34 per cent of titles experience a decline in lending. For around eight per cent of titles, there was no change during the project period.

There is no basis for claiming that the increase in lending is solely due to the pilot project. Additional comparative studies of general lending statistics compared with these extracts would be required in order to isolate the pilot project's effect on lending statistics.¹¹

However, there is a basis for assuming that the pilot project has had a contributing effect on the increase in lending: The increase is too high to be considered coincidental, and the lion's share of the growth in lending has occurred over the last year – almost two years after the Government announced its focus on the High North. There is also a correlation in time between exposure to search engines and media coverage of the pilot project on the one hand and the increase in lending of literature on the other.

The ratio between the ten most popular titles in the pilot project and lending of the same titles is shown in table 7.

Use of the service: library lending

⁹ Together with the example cited in the previous section, these are the only two instances, among those analysed, where a work has been read from cover to cover.

¹⁰ A comparable overview is available at <http://nb.no/nordomradene>

¹¹ The lending function providing a direct link from the search page to the lending service was in operation from 10 January 2008.

Title	No. of unique IP addresses	Lending 24.04.06-23.04.07	Lending 24.04.07-31.09.2008	Change in lending	%
The ten most accessed titles					
I nasjonalstatens tid, 1814-1940	222	65	95	30	46.15 %
Pan	159	1155	1310	155	13.42 %
Benoni	133	97	143	46	47.42 %
Norges religionshistorie	122	74	103	29	39.19 %
I kongenes tid, 900-1814	90	78	133	55	70.51 %
Historien om Nord-Norge	88	166	144	-22	-13.25 %
Norsk polarhistorie. Rikdommene	86	Not in the material	Not in the material	Not in the material	Not in the material
Knut Hamsun	83	6	12	6	100.00 %
Mot rikare mål å trå	80	2	0	-2	-100.00 %
Norge - et lite stykke verdenshistorie	77	107	117	10	9.35 %
TOTAL	1140	1750	2057	307	21.28 %
10 titles with the greatest increase in lending					
For kjærlighetens skyld: roman	14	281	1542	1261	448.75 %
Norsk-russisk ordbok	18	579	870	291	50.26 %
Den store boken om Norge	56	467	746	279	59.74 %
Blod, snørr og tårer	4	807	987	180	22.30 %
Norge på langs	15	430	594	164	38.14 %
Pan	172	1155	1310	155	13.42 %
Huset med den blinde glassveranda	30	1299	1452	153	11.78 %
Nådeløs villmark	5	884	1037	153	17.31 %
Norges bunader og samiske folkedrakter	15	501	639	138	27.54 %
90 dager på loffen i Børgefjell	3	416	522	106	25.48 %
TOTAL		6819	9699	2880	29.55 %¹²

Table 7: Use of the service: library lending

To a large extent, there has also been an increase in lending of books that are also available in the pilot project, but there seems to be no direct correlation between the number of visits paid to the titles and actual lending of the same titles. The increase in lending for these 10 titles is relatively large, however, with an increase of just over 21 per cent.

As can also be seen from table 7, the perspective has also been reversed in order to see how the ten titles displaying the greatest increase in library lending fare in relation to visits in the High North project. The figures show that no direct correlation can be documented between the titles with the greatest increase in lending and the number of unique visits to the same titles on the website. The increase in lending for these ten titles is relatively large, however, on average almost 30 per cent.

Evaluation

It cannot be ruled out that the government's focus on the High North as a political arena may have resulted in a general increase in interest in High North material. One tentative conclusion that may be drawn from the figures above is nevertheless that the pilot project's existence has contributed to an increase in lending of the literature covered by the project, but that there is not necessarily any clear, direct correlation between actual

¹² This figure does not include lending figures for the first work in the table. The reason for this is that there is reason to believe that the base material contains an error.

visits to the website and the subsequent lending of the same titles in the libraries.

3.9 Use of the service : sales

Description of the analysis

For the two periods forming the basis for the analysis of lending statistics, figures have also been obtained from publishing houses for the sale of High-North titles (see table 8). The service's sales portal was launched on 19 June 2008.

Results

Use of the service: sales					
Title	No. of unique IP addresses	Sales 24.04.06 - 23.04.07	Sales 24.04.07- 01.10.08	Change	%
The service's ten most accessed titles					
I nasjonalstatens tid, 1814-1940	222	664	14	-650	-97.98 %
Pan	159	38	265	227	597.37 %
Benoni	133	17	65	48	282.35 %
Norges religionshistorie	122	68	29	-39	-57.35 %
I kongenes tid, 900-1814	90	661	12	-649	98.18 %
Historien om Nord-Norge	88	234	1044	810	346.15 %
Norsk polarhistorie - Rikdommene	86	18	55	37	205.56 %
Knut Hamsun	83	0	0	0	0.00 %
Mot rikare mål å trå Norge - et lite stykke verdenshistorie	80	12	9	-3	-25.00 %
TOTAL	1140	2416	2199	-217	-4.70 %
The ten 10 titles loaned with the highest rate of growth					
	Change in lending				
For kjærlighetens skyld: roman	1261	60	250	190	316.67 %
Norsk-russisk ordbok	291	235	972	737	313.62 %
Den store boken om Norge	279	480	56	-424	-88.33 %
Blod, snørr og tårer	180	391	320	-71	-18.16 %
Norge på langs	164	19110	12740	-6370	-33.33 %
Pan	155	38	265	227	597.37 %
Huset med den blinde glassveranda	153	464	1710	1246	268.53 %
Nådeløs villmark	153	11640	7760	-3880	-33.33 %
Norges bunader og samiske folkedrakter	138	362	522	160	44.20 %
90 dager på loffen i Børgefjell	106	4320	2880	-1440	-33.33 %
TOTAL	2880	37100	27475	-9625	-14.91 %

Table 8: Use of the service: sales

Evaluation

There are great differences between the individual titles and it is difficult to draw any clear conclusions based on these figures. If one studies each title's year of publication and compares this with well-known "rules of the thumb" regarding general sales curves for books, the sales figures submitted primarily seem to be an expression of where in the general sales curve the title is located, whether revisions/republications have been made

during the course of the period in question and/or if the publishers' have initiated other market activities related to the individual titles. Conversations with publishing staff about the individual titles would appear to strengthen this impression.

If one looks at the lending figures and sales figures in context, one may ask oneself why the general increase in lending is not reflected in the sales figures. All the works studied are on sale in the normal way. Is it such that the increase in lending leads to a drop in sales? This question can neither be answered in the affirmative nor in the negative based on the figures available here. However, one cannot "isolate" the effect of the web-based service on sales figures. Since sales trends differ so much from title to title, it is difficult to point out the effect of the web-based service on the sale of individual titles. It seems more certain that the effect on sales is generally limited. There thus appear to be grounds to claim that the existence of the web-based service has had no significant effect on sales figures – neither a positive nor a negative effect. On the other hand, the sales function incorporated in the service has only been in operation for the last few months of the project period.

3.10 The survey

Description of the analysis

An analysis of the log data provides a good basis for describing the behaviour of users of the High North service, but it gives no good answer to questions such as who are its users, nor about their motivation and experiences as users. For this reason, a survey was conducted using the web-based tool QuestBack. The survey was conducted in the second half of October 2008.

The High North project largely concerns digital resources that are searchable from NBdigital and are closely integrated with the other services offered by the National Library of Norway. For this reason, it is not sufficient to direct the survey at those who access the web portal <http://www.nb.no/nordomradene/>, e.g. by using a pop-up window containing a questionnaire. The project group therefore chose an e-mail approach, picking two different lists of addressees:

- Subscribers to the National Library of Norway's newsletter, encompassing 633 individuals, described below as the U 1 list. One reminder was sent to this group.
- Membership lists from the Norwegian Non-fiction Writers and Translators Association, the Norwegian Critics' Association and LINO, totalling 4223 individuals, called the U 2 list below. One mailing. The figures stated here have been adjusted for returned e-mails due to incorrect e-mail addresses, absences etc.

The survey consisted of 11 multiple-choice questions.

Results

Twenty-six per cent of users from the U 1 list responded (163 of 633 addressees), while 22 per cent of users from the U 2 list responded (929 of 4223 addressees).¹³

The questions posed in the survey fall into three categories:

- Information about users
- The users' familiarity with the High North project
- The users' experiences of the service

The responses are given in table 8.

¹³ From a statistical point of view, the number of respondents is low, but as far as web-based surveys are concerned, the response rate is relatively high.

The survey								
User information								
Gender	Male	Female						N¹⁴
U 1	46.0 %	54.0 %						163
U 2	66.6 %	33.4 %						1058
Age	0–19 years	20–39 years	40–59 years	60–79 years	80–99 years	N		
U 1	0.0 %	16.0 %	41.1 %	40.5 %	2.5 %	163		
U 2	0.0 %	6.3 %	48.4 %	43.2 %	2.0 %	1058		
Sector	Private	Public	In education	Not in employment	N			
U 1	26.4	54.6	4.9	14.10 %	163			
U 2	35.2	52.5	0.6	11.70 %	1058			
Profession	Press/media	ICT	Library	Student	Researcher	Other	N	
U 1	9.8 %	0.0 %	8.6 %	4.9 %	20.2 %	56.4 %	163	
U 2	13.9 %	0.9 %	1.4 %	0.1 %	39.5 %	44.3 %	1058	
Region	Northern Norway	Central Norway	Western Norway	Southern Norway	Southeastern Norway	N		
U 1	3.1 %	2.5 %	5.5 %	1.8 %	87.1 % ¹⁵	163		
U 2	7.5 %	10.0 %	16.7 %	4.3 %	61.5 %	1058		
Users' familiarity with the High North project								
Familiar with the project	Yes	No						N
U 1	20.2 %	79.8 %						163
U 2	11.0 %	89.0 %						1058
How did you find out about the project	Homepage of nb.no	Search in NB	Search engines	Other websites	NB 21	Newsletter	Media	Other
U 1	48.5 %	3.0 %	3.0 %	6.1 %	3.0 %	18.2 %	6.1 %	12.1 %
U 2	19.6 %	5.9 %	13.7 %	4.9 %	3.9 %	2.0 %	15.7 %	34.3 %
Users' experiences of the service								
Did the user find any interesting material in the service (1 = to a limited extent, 6 = to a great extent)?								
	1	2	3	4	5	6	N	
U 1	3,0 %	0,0 %	15,2 %	39,4 %	36,4 %	6,1 %	33	
U 2	5,8 %	12,6 %	7,8 %	15,5 %	35,9 %	22,3 %	119	
What type of material did the user look at?								
	Books	Journals	Films	Radio	Photographs	N		
U 1	80,6 %	39,8 %	14,6 %	5,8 %	37,9 %	33		
U 2	69,7 %	42,4 %	15,2 %	15,2 %	54,5 %	119		
What was the user's experience of reading books and journals online (1 =very bad, 6 = very good)?								
	1	2	3	4	5	6	N	
U 1	0.0 %	14.8 %	25.9 %	25.9 %	29.6 %	3.7 %	27	
U 2	5.6 %	5.6 %	35.6 %	30.0 %	20.0 %	3.3 %	105	

Table 8: The survey

¹⁴ Number of respondents.

¹⁵ The reason why the vast majority of respondents on the U1 list come from southeastern Norway is because the newsletter largely mentions events that take place on the National Library of Norway's premises in Henrik Ibsensgt. 110 in Oslo, and for this reason most addressees on the list come from Oslo and its environs.

Both the responses and the comments are predominantly positive, both in respect of the High North project and the service itself. Most users found interesting material. Books and journals were accessed most. Reading online is not perceived as being a problem. Some respondents would like the opportunity to print out interesting material. The group aged between 60 and 79 years comprises 40.5 per cent of users on the U 1 list and 43.2 per cent av users on the U 2 list. The main bulk of users is between 40 and 80 years of age. More than two per cent are aged over 80.

Evaluation

Even though the composition of the respondents cannot be said to comprise a representative sample of the population and the numbers are generally low, the responses to the survey nevertheless give a good indication of how users who visit the National Library of Norway and similar services perceive the High North project. In the opinion of the project management, the survey – wherever it is natural – supports the results of other analyses presented in this report.

One flaw in the survey is the fact that there are no respondents in the group aged between 0 and 19 years. The reason for this is that that representatives for this age group can hardly be found in the address lists that were used. It can otherwise be seen that almost twice as many individuals from the U 1 list were already familiar with the service; almost half of this group had become familiar with the website via the homepage of the National Library of Norway. The respondents from the U 2 list, on the other hand, arrived via search engines and the media. As regards this latter group, it may be mentioned that after good media coverage of the project in May 2008, visitors to the service almost tripled compared with the months before and after.

4. Overall evaluation

- The fact that 89 per cent of books and 85 per cent of journal articles have been visited documents that users have found the service interesting. The project management has been unable to find surveys elsewhere that may be fully compared with this survey, in order to refer to a relevant reference. One example dealing with user frequencies: "Nearly three-quarters of sessions saw content viewed ..."¹⁶ The investigation cited analyses a project at University College London, and deals with students' use of the curriculum.
- The fact that 65 per cent of titles in the service that were accessed were published no later than 11 years ago may signalise that the internet is an effective channel for distributing literature, particularly titles that are difficult to access in other ways. It can also be seen that there is a good distribution among the subject areas in which there is an interest. The fact that such titles are made available also increases demand. Otherwise usage frequency increases in accordance with the time the service is operative: time generates usage. And credit should go to smaller markets or niches. This phenomenon is generally called the Long Tail, a term that was first launched by the American writer Chris Andersson. The theory is that the vast majority of products (e.g. books) that are not immediate bestsellers have virtually the same value as those titles where demand is greatest, simply because they are numerous and sell over a long period of time. Andersson says: "This is not just a quantitative change, but a qualitative one, too. Bringing niches within reach reveals latent demand for non-commercial content. Then, demand shifts towards the niches [...], creating a positive feedback loop that will transform entire industries – and the culture – for

¹⁶ David Nicholas et al.: "SuperBook", in the programme accompanying the "Online Information 2007" conference, London 4-6 December 2007, p. 50.

decades to come".¹⁷ The figures in this present analysis would appear to display a clear "tail effect".

- One deficiency of the survey is that it does not detect young users. In spite of its limited representation of the Norwegian population as a whole, the survey nevertheless gives a good indication of the actual users' usage patterns, and supports to a certain degree the results found in other analyses.
- The footprint analysis documents that the ten most active users have a highly variable usage pattern. Some access a large number of titles; others make do with a small handful of titles. The fewer titles a user accesses, the more pages he or she reads. And one striking feature is that users generally access text pages. On the other hand, the majority of users appear to use the service in the same way one uses a library, namely as a place where one finds something interesting or something one needs. To a lesser extent, the website is used as a "reading room", even though one can find exceptions to this rule (among other things, two of the fiction titles analysed were most probably read in their entirety online).
- Is it possible to document that the existence of the High North project has had an impact on library lending? The project management's tentative conclusion is that during the project period there has been an increase in lending of the literature included in the project. It is not possible to demonstrate a direct correlation between the number of hits on the website and the ensuing number of loans of the same titles from libraries.
- Has the existence of the High North project affected the sale of titles included in the project? It does not look like the existence of the service has had any noticeable effect on sales – neither a positive nor a negative effect. On the other hand, the service's sales function was only operative for the last months of the sales period.
- The number of users and the number of pages viewed increased considerably when search engines were given access to the service's metadata in January 2008. This tallies with the findings of studies in other countries. One example: "Those people accessing via search engine were most likely to record more views in a session and were more likely to view text pages."¹⁸
- The surveys referred to above (notes 7, 16 and 18) analyse two digitalisation projects at academic institutions in the United Kingdom and the USA. The project management has studied these analyses but has found that they are only suited as a frame of reference to a limited extent; as indicated by differences between target groups, various lengths of the project periods and deviating criteria of analysis. On the other hand, one can still ascertain that the tendency and usage patterns documented also substantiates to some extent the results in the present analysis.

5. The road ahead

One of the goals of the pilot project has been to gain experiences that can provide impulses for continued collaboration on the dissemination of copyrighted material in digital form.

The study presented here shows good usage of the service. The material requested is significantly diverse in the preferred subject areas and the age distribution is interesting in relation to year of publication. The project management has resolved to extend the project period to 31 December 2009 in order to further assure the quality of the results obtained.

¹⁷ See Chris Andersson: *The Long Tail. How Endless Choice is Creating Unlimited Demand*, Random House 2006, s. 26. The quote included at the beginning of this report is taken from Chris Andersson's article "Google and the Long Tail of Time", http://thelongtail.com/the-long-tail/2006/04/google_and_the_.html, s. 1.

¹⁸ Nichols, et al.: "SuperBook", op.cit., p. 54. See also David Nichols et al.: "What does usage data tell us about users?", op.cit. p. 84.

In parallel with the desire to streamline the pilot project, the project steering committee has considered the opportunities of extending their collaboration. Specifically, the parties have jointly begun to plan a new and more comprehensive project which is due to be launched in April 2009. The working title of the new project is "The Norwegian Bookshelf" – in which all Norwegian books published in the three decades 1790, 1890 and 1990 will be made available in digital form.