Nichesourcing: a case study for pluvial data digitization for the Sahel

Binyam Mersea Tesfa
Information Sciences
VU University Amsterdam
Amsterdam, Netherlands
binmer2@gmail.com

ABSTRACT

In this paper, we describe a Nichesourcing case research for digitizing pluvial data from the Sahel region in Africa. Crowdsourcing is a popular approach to digitize documents on the web; however it also has limitations with respect to maintaining highly motivated participants and handling complex user tasks. In order to deal with the limitations of Crowdsourcing and based on 'the task and its complexity', 'the product' and 'the resource pool' we identified Nichesourcing as suitable approach for this project. Nichesourcing is a variation of Crowdsourcing that targets specific niche groups instead of targeting the general crowd. We developed and published a Nichesourcing application on the web and evaluated its success in terms of attracting dedicated participants and digitizing considerable amount of digital data. With one week release of our Nichesourcing application, we have produced more than 5000 cells of structured digitized pluvial data. We also found that the anticipated niche (people with African affiliation) dedicatedly participated in the digitization. However, the project run only on a small scale and it will be necessary to perform further research to proof that Nichesourcing suits this type of projects.

Categories and Subject Descriptors

H. 4.2 [**Information Systems Applications**]: Types of Systems

General Terms

Theory, Design

Keywords

Nichesourcing, Crowdsourcing, Incentives,

1. INTRODUCTION

The Web has brought knowledge sharing and socialization to a massive scale and speed that was not seen before. With the rapid growth of the global Internet penetration the world is sharing and utilizing data progressively through the web. However, the penetration of Internet and utilization of the web varies in different part of the world. Not yet everyone on this planet takes advantage of the web's important benefits such as linking and accessing information: still 4.5 billion of people do not have access to the [19]. There are also large variations in the Internet penetration and utilization rates among different regions of the world. According to a report from Internet World Stats [11] the Internet penetration rate in North America is 78, 9%, the highest in the world, while Africa has the lowest Internet penetration rate of 13.5% [11].

Several initiatives have been taken to enhance knowledge sharing through the web in different parts of the world. One of those initiatives is the Web alliance for Regreening in Africa (W4RA), which aims to extend the web's benefits of the knowledge society and economy to people in rural communities in Africa [19]. The World Wide Web is still not accessible to the majority of people in Africa mainly due to the absence of proper infrastructure. Alternative or complementary technologies can however be used to extend the web's accessibility within the local context in Africa. For example, through the W4RA a voice- and web based market information system 'RadioMarché' is deployed aiming to stimulate agricultural trade in Sahel countries in Africa [3]. RadioMarché (RM) is being developed within the context of the VOICES (VOIce-based Community cEntric mobile Services) project. In a nutshell, RM produces a digital/computerized version of the local market information, links the market information with other relevant data and makes it accessible via several formats (including mobile telephone voice in a local dialect) for everyone connected to the Web. In such a way, RM uses mobile-voice interface to overcome interfacing and infrastructural issues in the region.

Despite the infrastructural issues in Africa, benefits of the web can further be extended to support initiatives in Africa. In this thesis research, we aimed to develop an application on the web that taps into the power of a motivated crowd to digitize handwritten pluvial data from Sahel region in Africa.

We compared whether the general Crowdsourcing (popular approach to organize web users to perform tasks) or the more specific Nichesourcing suits this project. Based on 'the task and its complexity', 'the product' and 'the resource pool' we anticipate that Nichesourcing suits this project by dealing with the limitation of Crowdsourcing in terms of (i) the quality of the crowd's input/final result, and (ii) the ability to solve complex tasks. We also anticipate 'people with Africa related background' is the right niche group to target based on the extra effort and motivation the digitization requires. In order to assess our claims, we developed and published a Nichesourcing application on the web and discuss our approach and findings in this paper.

This paper continues with providing the theoretical background of this research project in chapter 2. Subsequently chapter 3 describes the use case of this study, the digitalization of pluvial data from the Sahel region, and the niche group of people with an African background. Chapter 4 illustrates the Crowdsourcing application that was developed for this project. In chapter 5, the evaluation measures are described followed by our findings and conclusion in chapter 6 and 7 consecutively.

Conference'10, Month 1–2, 2010, City, State, Country.

2. THEORETICAL BACKGROUND

A generally popular approach to organize web users to perform tasks is to use human-based computation systems such as Crowdsourcing [13]. Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call [7]. Crowdsourcing has been used for a variety of tasks, amongst others the digitization of data. From 2007 to 2009 alone approximately 400 million people have helped digitalize more than five billion words [8]. Several types of digitization tasks have been performed through Crowdsourcing applications. A typical example is the "Old Weather" initiative [21]. The "Old Weather" initiative launched a Crowdsourcing application that aimed to collect data on temperatures from historical ship logs. Through the "Old Weather" Crowdsourcing application 202,904 pages have been transcribed within three months of its launch in October 2010 to December 2010 [12].

Although Crowdsourcing has been successfully applied to many areas, there are also difficulties with the design and prototyping of the systems. One source of uncertainty during design is the behavior of the crowd at different levels of scale. A second source of uncertainty during design is the crowd's response to differing incentive functions [18]. Furthermore, crowd collaboration has critical challenges of finding sufficient knowledgeable and loyal users and also maintaining a reasonable level of quality [12]. In some Crowdsourcing projects a malware and spam of up to 95% [5] had been found and such drawback makes the quality (and consequently the success) of some Crowdsourcing project results questionable.

A recently evolving variation of Crowdsourcing, which is aimed to tackle the drawbacks of Crowdsourcing in terms of (i) the quality of the crowd's input/final result, and (ii) the ability to solve complex tasks, is Nichesourcing. Nichesourcing is a specific type of Crowdsourcing where complex tasks are distributed amongst a small crowd of experts (for example art enthusiasts or African expats) rather than the "faceless" crowd [4]. A niche is gathered from either distributed experts on a specific topic or from an existing network cantered around the same culture, location or topic. In both cases the members of the niche group have domain knowledge and/or an intrinsic motivation to contribute and provide high quality results.

To explain the major differences between Crowdsourcing and Nichesourcing, three main aspects could be considered from a process-centric perspective [4]. Based on these three aspects, namely (a) the task and its complexity (b) the product and (c) the resource pool, we can anticipate whether Crowdsourcing or Nichesourcing is more suitable for a particular project.

The first identified aspect to be considered in determining whether to target a random or specific crowd is (a) the task and its complexity. In both systems large tasks are divided into smaller, atomic tasks. Crowdsourcing mainly focuses on repeatable atomic tasks which can be performed by a random crowd. Nichesourcing is however appropriate for more result oriented atomic tasks that may need a specific level of effort and/or knowledge; and which can be performed by a particular crowd with the anticipated motivation and/or the particular knowledge.

Another identified aspect for consideration in assessing the suitability of Crowdsourcing vs. Nichesourcing is (b) the product that is targeted as a result of the task. Generally Crowdsourcing focuses on quantity, whereas Nichesourcing focuses more on the quality of results. Although Crowdsourcing will usually reach more crowd/participants than Nichesourcing, the latter might be more suitable for tasks that require more control over the quality of the results.

Finally, (c) the resource pool of participants needed to perform the tasks is also identified as an aspect to be considered in anticipating the most suitable crowd. Quinn et al. [13] have identified five types of motivation for participants in human-based computation applications: payment, altruism, enjoyment, reputation and implicit work. In Crowdsourcing, a key measure of success is the size of the crowd, hence building and maintaining a dedicated crowd by using payment, enjoyment and reputation (public recognition for efforts) as incentives could be most effective. In Nichesourcing, where the quality of the results is a key measure of success, motivating participants by appealing to their altruism (wish to help or 'do good') could be more effective.

This research project aims to explore the usage of Nichesourcing for the digitization of African agricultural and environmental data, in order to make the data available for re-greening initiatives. For this purpose a Nichesourcing application has been developed and published on the web and made accessible through Facebook, the largest social network on the web. Through contributions from people from all over the world more than 5,000 cells of digital rainfall and harvest records were digitized in a period of one week time. From this experience conclusions and lessons learnt are drawn.

3. USE CASE: Pluvial data from Sahel

3.1 Description of the case

A lot of potentially useful data in Africa is still locked on written paper. Governments in the Sahel region have recorded thousands of handwritten pages of data about rainfall, harvests and floods in their region. If this data would be made digitally available and easily accessible, and combined with other data, it could be used to support decisions in re-greening initiatives both locally and globally.

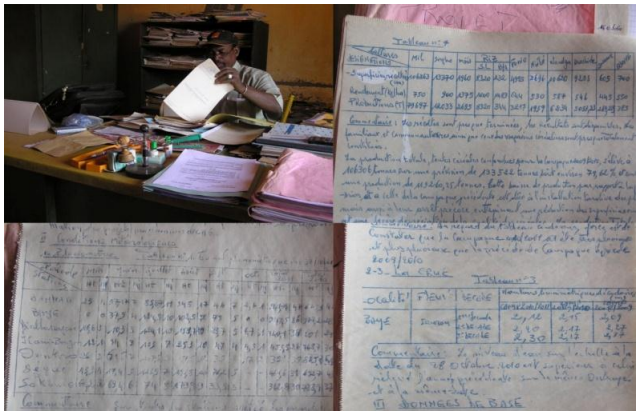


Figure 1, Sample hand written pluvial data forms from Mali.

Digitization of handwritten data in the Sahel region is subject to several challenges. The limited infrastructure in the region hampers the use of automatic digitization techniques. Manually

digitizing thousands of pages of records would also be very time taking for the usually understaffed regional governments.

3.2 Sample Pluvial records

For the case analysis we made use of twenty-eight sample digital images (captured by digital camera) of handwritten pluvial records from the Sahel region. We obtained these pictures through Wendelien Tuyp, an expert working with Africa Regreening Initiatives.

These sample handwritten documents consist of yearly details of pluvial records from 2008 to 2011 in French language. Based on the structure of the document and further discussion with Wendelien Tuyp and other French speakers we found out that the sample documents contain flood, harvest and rainfall records. Table1 below summarizes the structure and content of the documents.

Table 1, Review of total sample pluvial records obtained

Record Structure	Enclosed Information	Rows	Column	Count
Textual	Introduction, Comments, and Review	5	1	7
Tabular	Flood	3	5	4
Tabular	Harvest	3	14	3
Tabular	Rainfall	7	17	4

3.3 Crowdsourcing versus Nichesourcing

We examined whether Crowdsourcing or Nichesourcing suits this project according to the basic framework provided to compare their three main aspects. As described in the chapter 2, de Boer et al. [4] have identified three aspects that could be considered from a process-centric perspective for anticipating whether Crowdsourcing or Nichesourcing is more suitable for a particular project: (a) the task and its complexity (b) the product and (c) the resource pool.

a) The task and its complexity

Based on the sample documents and the discussions held, we found out that the records contain inconsistent textual information: while there is textual information included in the records of some years (introduction, comments and market reviews), there is no corresponding textual information included in the records of other years. A reasonable command of French is also required for the digitization of the textual parts of the records since some of the handwritten contents are very hard to read. In addition, familiarity with names of local towns and villages, although not necessary to perform the tasks, could increase the efficiency of the transcription.

Furthermore, the yearly rainfall, flood and harvest records vary in structure and detail. For example, as shown in figure 2, the numerical flood records from 2008 consist of 2 columns of data, while the same records from 2009 consist of one column of data, and the same records from 2011 consist of 3 columns of data.

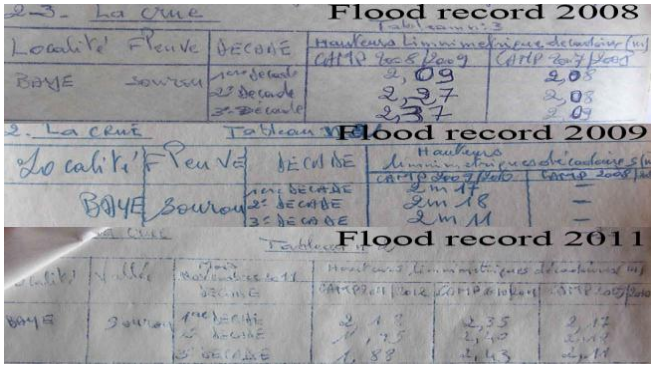


Figure 2, Inconsistent columns in flood records from the years 2008, 2009 and 2011

The division of the records into very small atomic parts is a solution to make the task suitable for Crowdsourcing; however this will complicate the reconstruction of the whole document out of the atomic parts. For example, the flood records (see figure 2 for sample) has about 15 cells of records on average. It is possible to divide the flood records in 15 atomic parts in order to fit ordinary crowd who might prefer to digitize a single cell of record. However, it would require more complex design to effectively reconstruct the row-column relationship in between the 15 parts.

Some of the pictures of the handwritten records are difficult to read and partially unreadable, hence it needs extra effort to digitize as accurately as possible. Knowing French language and familiarity with the names of the villages will also highly benefit the digitization of the textual data.

Based on the complexity of the task (extra effort and knowledge it require), we anticipate that Nichesourcing would be more suitable for this project than Crowdsourcing.

b) The product

The end product, i.e., the digital version of the records, will be especially useful when aggregated over multiple regions for statistical analysis supporting decisions in re-greening initiatives. Since the records are therefore interrelated, wrong transcription of a single record may decrease the accuracy of the aggregated result. For example, a wrong transcription of rainfall record of a specific Month of the year on a specific location (such as Bankass October, 2011) will bias the aggregated yearly rainfall record. Consequently, quality is a key success factor for the product, and therefore we anticipate that Nichesourcing is more suitable than Crowdsourcing for this project.

c) The resource pool

Due to the effort and specific background information that is required, a particular niche group which possesses the specific knowledge and motivation would be more suitable than a general crowd. As described in a) above, participants who know French language or who are familiar with the names of the villages are more suitable to digitize the textual records. Highly motivated participants are required to participate in order to perform the fairly complex tasks with reasonable quality. A niche group, community of practice [20], with native curtsey to actively participate would best fit the resource pool need of this project.

Since the digitization of these records mainly benefits re-greening initiatives in Africa, a specific niche group (people with African

background and/or environmentalists) is expected to keenly participate.

3.4 The niche group: people with Africa related background

The niche group that we targeted in this project consists of people with an Africa related background, including Africans living in Africa, African Diaspora and non-Africans who are living or have lived in Africa. Due to an affiliation with African issues, people with an African background would be higher motivated to participate in this project than members of a general crowd. In addition, a relatively large part of this group is expected to have a good command of French.

Although most Africans living in Africa are still not connected to the web mainly due to a lack of sufficient infrastructure, there is a steady increase in Internet penetration and user rates on the continent. Especially African youth is increasingly using the Internet and engaging in social media [1].

There are also millions of people of African origin living in different and better connected parts of the world. The African Diaspora is defined by the African Union (AU) as "consisting of peoples of African origin living outside the continent, irrespective of their citizenship and nationality, and who are willing to contribute to the development of the continent and the building of the African Union". The AU estimates that the African Diaspora consists of approximately: 112.6 million people in South America; 39.2 million people in North America; 13.5 million people in the Caribbean; and about 3.5 million people in Europe [17]. Even though they live in different parts of the world, members of the Diaspora are usually very much affiliated with local issues in their region of origin.

A third subgroup of people with an African background consists of non-Africans who are living in Africa or have done so in the past, for instance expats from different parts of the world.

In this research, we investigated if the anticipated niche (participants with an Africa related background) participated more actively than other participants.

4. THE REGREENING AFRICA APP

4.1 General Introduction

We designed, developed and published a Nichesourcing application, named Regreening Africa App. The Regreening Africa App is designed based on the basic structure of the sample pluvial records, obtained from Mali, in a way that it can be re-used for digitization of any pluvial data confirming to the same structure of documents. We develop the system with popular open source, cross-platform compatible standard technologies (such as PHP, mysql and Drupal) to insure the re-usability of the system as much as possible. The Regreening Africa App was hosted on a private server of the researchers [15], for one week release as part of this study. It was also made accessible through the biggest social media site Facebook as Regreening app [16] in order to enhance its' accessibility. Using a Facebook plug-in, the system allowed easy login and sharing options for Facebook users.

Figure 3, Facebook access to Regreening Africa App: showing how to reach the app

In a nutshell, the application (I) uploads documents to be digitized (II) provides digitization tasks for participants (distribute digitization) and (III) stores the digitized version of the documents. The system also provides background information on the project, textual and pictorial explanation on how to participate, and answers to frequently asked questions (FAQ).

In order to make participation as easy as possible, logging in was not made mandatory for performing digitization task through the system. However, participants were recommended to log in with their username (i.e. using their Facebook login name or by creating a username through the system) while transcribing. Points were also awarded for logged digitization as an incentive (see Sect. 4.2 for more details).

I. Uploading documents to be digitized

An authenticated user or domain expert can upload digitization tasks (jpg images of the handwritten records) to be digitized. To upload a task, the expert needs to make /arrange/ a low resolution picture of a record first (For example: Table of Flood record 2008) and then upload the document and its basic detail through a back end interface as shown in figure 4 below.

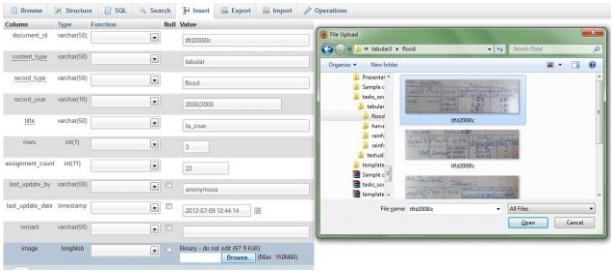


Figure 4, Uploading digitization tasks

II. Providing digitization tasks

Once records to be digitized are uploaded in the system, the application provides digitization tasks to its users. The digitization tasks are classified in five groups based on the average time (effort) the task requires to complete and the specific knowledge it may require (see Sect. 4.2 for more details). We made informed estimates of the average time each group of task took by performing the tasks with different pace and averaging the time it took to complete. We then name the tasks as Tree task, Garden task, Park task, Forest task and Grand Forest task to symbolically correspond the digitization to re-greening a tree, a garden, and a park, a Forest or a big forest consecutively.

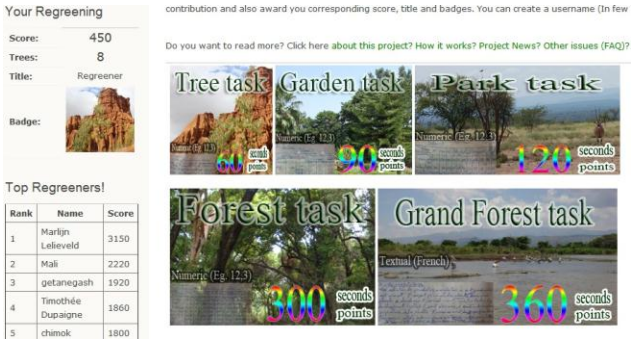


Figure 5, Regreening Africa App home page showing the 5 types of digitization tasks (main content), logged in user digitization points (Left bar top) and Top Regreeners list (Left bar bottom).

Participants select the type of task they would like to perform by clicking on the image of the task (or the name of the task from the top menu), after which the corresponding task is provided by the system (see figure 5). A picture of handwritten records to be digitized is shown to the participant, with a data entry form underneath. As shown in figure 6, the data entry form has a similar format as the data to be digitized so that users can easily copy the data from the top image into the form below.

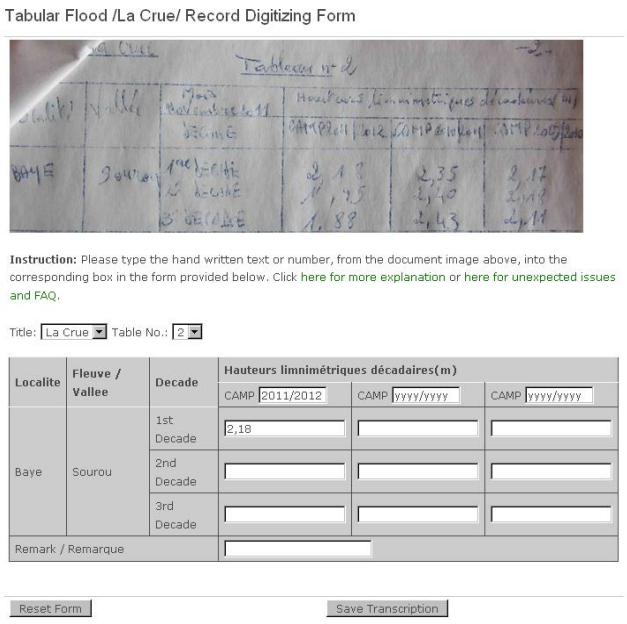


Figure 6, Digitization guide: Tabular flood data digitization form with guiding arrows

The application also awards a score, trees, a title and a badge to participants based on the effort (the time the task is estimated to take). For example, the smallest task (tree task), estimated to take on average 60 seconds to complete, is awarded with 60 points, while the most complex task which requires additional skill of French language (Forest task), estimated to take on average 360 seconds to complete, is awarded with 360 points. For logged in users, the system shows their aggregate Regreening score, number of trees, title and badge on the left bar of the website. In addition, a ranking of "Top Regreeners", mentioning the name and score of the 5 most active participants, is shown on the left side bar to further encourage participants. (See 4.2 for more explanation)

III. Storing the digitized version of the documents

The system stores each transcription/digitized version of the records and at the same time maintains the relational components. This way, easy and flexible access is possible to either a particular record or an aggregated value. For example, a rainfall record for a specific month, year and location could be accessed as well as the maximum rainfall record from all the digitized records could be easily computed. To this end we design a relational database that can store the records while keeping the relationship in between the records (Figure 7).

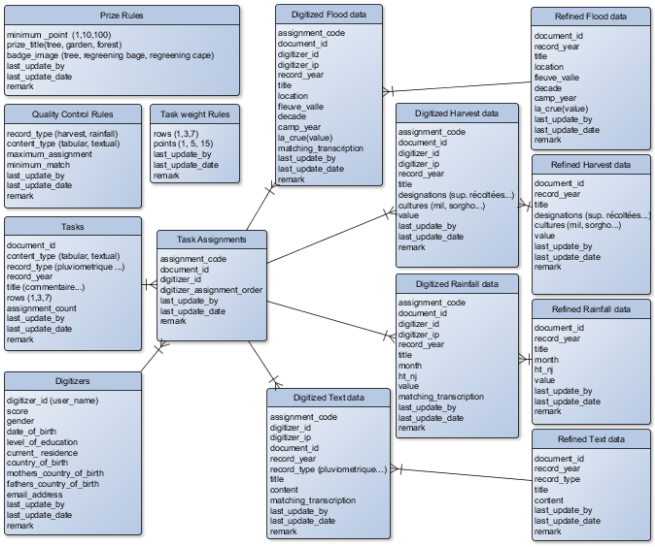


Figure 7, Relational database model: Regreening Africa App

4.2 The design and implementation decisions

The major decisions in designing this application involved the task decomposition and reward/motivation.

I. Task decomposition

As described above, the tasks were classified into five groups based on the time the tasks take to be completed. From a design perspective the tasks were however also categorized based on the content of the records (flood, rainfall and harvest) and the type of the records (numerical or textual). Table 2 below shows the differentiation between the tasks. Since the rainfall records are very long, they were split into two groups: Tree task (one row) and Forest task (7 rows).

Table 2: Task decomposition

Group of tasks	Content	Type of data	row	column	Time it take
Tree task	Rainfall	numerical	1	16	60 sec
Garden task	Flood	numerical	3	3	90 sec
Park task	Harvest	numerical	3	14	120 sec
Forest task	Rainfall	numerical	7	16	300 sec
Grand Forest task	Rainfall, flood, harvest	textual	1	1	360 sec

For each group of tasks a corresponding template was designed, in which users could type their transcriptions. The template that was designed for the textual data (Grand Forest task), included a selectable title list (introduction, commentaire, le Maraichage) and a data entry box for text (see figure 8). The template that was designed for the numerical data contains the general headings (which are common in the related records) and provides data entry boxes for the numerical values.

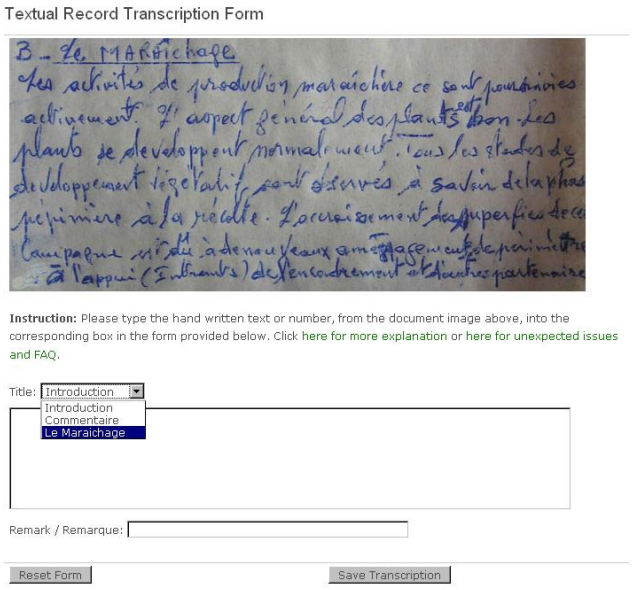


Figure 8, Textual record transcription form

Unlike Crowdsourcing tasks, the atomic tasks are fairly complex. The minimum numerical digitization task (Tree task) requires the effort of transcribing 16 cells of records which could have been divided to smaller part if we were addressing a general crowd. The textual digitization task also requires additional skill of French language.

Nonetheless, we anticipate that the intrinsic motivation and background knowledge of the niche will produce significant result. The level of recurrent participation of the niche could indicate the level of motivation they have.

II. Reward and motivation: Gamification

How to improve user engagement has become one of the essential challenges in designing this type of systems. The current trend of integrating game design and game thinking into non-game applications, the so called Gamification concept, suggested a new solution for better engaging with end users. Although in general the design and implementation of Gamification powered systems require considerable effort, these systems have a greater impact with regards to the ability to incentivize and influence people [9].

Figure 9: The Gamification loop [9]

Regreening Africa App Gamification incentives

We incorporated the main game mechanics used to gamify a system, including a point system, reward and badges as shown above in Figure 9. We did not use the entire framework due to time shortage, however we have incorporated 5 levels of incentives based on points of the aggregate digitization (time the tasks require to complete) as illustrated in table 3 below.

The badge images used for this release of the system are quickly selected from personal non-copyrighted pictures that the researcher captured in Africa. A better quality and more descriptive badge could be easily incorporated for future release.

Table 3: Regreening Africa App Gamification level design

Minimum point	Prize title	Badge image	Badge Description	Win condition
60	Regreener		A single tree in a dry land	At least 1 tasks completed
500	Regreening Hero		Attended garden with variety of plants	At least 2 tasks completed
1000	Regreening Super Hero		A natural park with a wild animal	At least 3 tasks completed
2000	Regreening King		A green natural forest	At least 6 tasks completed
5000	Regreening Lord		A big forest with big river	At least 13 tasks completed

5. THE EVALUATION

5.1 Evaluated aspects

This research project aims to investigate if Nichesourcing is suitable for the digitization of pluvial data from Africa, and if people with Africa related background is an appropriate niche group to be targeted. For this purpose we developed a Nichesourcing application that was published on the web and made accessible through Facebook.

We couldn't get more pluvial data to be digitized, than the very small (18) number of sample records, due to the outbreak of war in Mali and the limited time we had. Consequently, we only used the 18 tasks grouped in 5 types of tasks. Due to the very limited number of digitization tasks in the system, a digitizer who perform one type of task for 5 times on average (such as doing 5 times tree task) will end up doing the task s/he already did. Therefore, we decided to run the system online for only one week. This way we aim to decrease the chance of users' dissatisfaction that may arise from doing the same task repeatedly.

To evaluate our approach and application we assess the following aspects:

- The quality of the provided result. Highly similar digitization by different digitizers indicates better quality.
- The number of people that made use of the application and the amount of data that was digitized by these users. Producing more data with few participants may indicate reaching motivated niche.
- The background of the participants that participated: Africans living in Africa, African Diaspora, non-Africans who lived in Africa, participants without an African background
- How often people participated and whether there are differences in the frequency of participation between the above mentioned groups
- Indications for possible motivation of participants and differences in motivation between the identified groups of participants
- The usability of the application

5.2 Setup of the experment

The Regreening Africa App was published online [15] for a period of one week. In addition, the app was made accessible through Facebook as Regreening App [16]. The app on Facebook mainly provides an iframe that loads the website through Facebook theme. The entire system (including user management and storing the digitized version of the records) runs from the main website.

An open invitation for participation (see appendix 1) was forwarded by email and through Facebook to and through the personal and professional contacts of the researcher. The application was deployed for one week and general details of users' activity were recorded for evaluation purposes. These details include: the participants' ID or anonymous-user tag if not logged in, the participants' digitization assignment order/reputation and the digitization task s/he was assigned to.

After the one week deployment of the application, participants were requested by email and on the website of the application to fill in an online questionnaire (see appendix 2). The survey was carried out in order to assess the usability of the system and to receive information about the participants' background and motivation.

The questionnaire first asked participants about their gender, age, education level, their (African) background, and whether they visited the Regreening system and created and used a login account. Subsequently five questions were asked to assess what motivated participants to do the digitization tasks. These questions are mainly based on the five types of motivation for participants in human-based computation applications: payment, altruism, enjoyment, reputation and implicit work [13]. One more question (question 6) is also added to get indication about the influence of knowing the researchers on the participation. Participants were asked to indicate to that extent they agreed with the following statements by making use of a five point scale (1= strongly disagree, 5=strongly agree).

1. I was motivated to participate in projects benefiting Africa
2. I was motivated to participate in re-greening initiatives

3. I was motivated to do anything good with the time I had
4. I was motivated because it was fun and enjoyable
5. I was motivated because I received points and recognition
6. I was motivated because I personally know part of the researchers and/or participants involved

The last component of the questionnaire consisted of the system usability scale (SUS) referenced in more than 600 publication [14]. The SUS is a ten item questionnaire with five response options (1=strongly disagree, 5 = strongly agree) containing the following questions.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

5.3 Result

Analysis of the outcomes of the survey and the data gathered during the one week release of the application yields the following results.

✓ Quality of the result

Although detailed analysis on the quality of the result is not performed (due to time shortage), a brief look at the results show that there is a very high similarity /agreement/ in the digitized versions of most cells (see Appendix 3 for more). The only considerable inconsistency in the result is on the use of digit separators comma versus dot (such as 206, 9 and 206.9).

✓ Amount of digitized data and number of participants

Within the one week release of the application more than 5,000 cells of records were digitized and a total of 266 digitization tasks were completed. In order to make participation as easy as possible, logging in was not made mandatory to performing a digitization task. Due to this the total number of participants is however unknown. Almost half, 48% (127), of the digitization tasks were performed by participants who did not log in while digitizing. The other half, 52% (139), of the tasks were performed by 22 participants/ registered users who performed the digitization tasks while being logged in with their username.

Although the number of participants is very low the amount of complex tasks completed with a reasonable quality of the results shows the participants dedication and this corresponds to the expectations found in Nichesourcing.

✓ Africa related background of participants

20 respondents who vary in age, gender and level of education completed the online survey questionnaire. A very large majority of the respondents (90%) filled in that they have an Africa related background. As shown in figure 10.

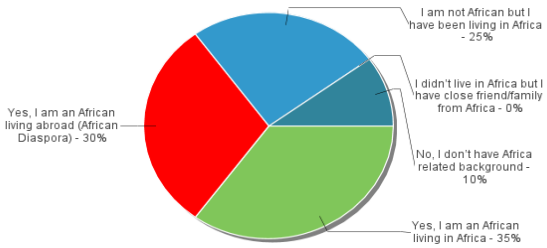


Figure 10, Questionnaire respondents' background

✓ Frequency of participation

A large majority, 87%, of the registered participants performed tasks recurrently. From the registered users 13.6% performed 1 digitization task, 50% performed 2 to 5 tasks and the remaining 36.4% performed 6 to 24 digitization tasks (see figure 11).

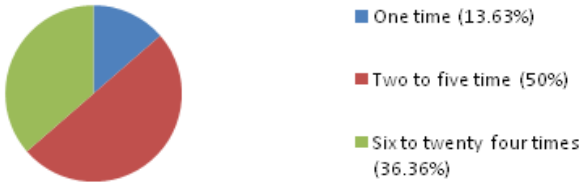


Figure 11, Digitizers' participation recurrence

✓ Motivation of participants

A very large majority of the respondents (70%) filled in that they are motivated in 'projects benefiting Africa'. As shown in figure 12, more respondents filled in that they are motivated in 'projects benefiting Africa' and 'doing good' more than 'earning points' or 'enjoyment'.

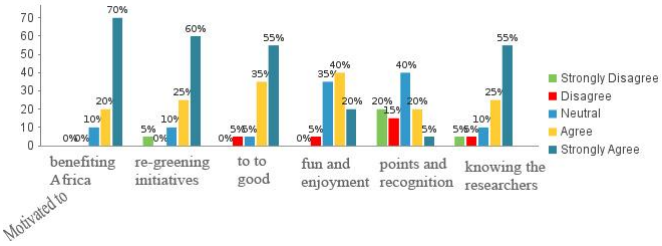


Figure 12, Digitizers' response about their motivation to participate

The review of the questioner on the motivation of participants indicates that participants were more motivated by altruism (wish to help or 'do good') which corresponds to correspond to our expectation of Nichesourcing.

- ✓ Usability of the system

The system usability (SUS) score is 71.87, which can be interpreted as the system shows above average perceived usability [10].

The respondents were also asked whether they created and used a login account. One third (33%) of the respondents indicated that they created and used a login account, while another 10% stated that they did some tasks as anonymous user first and then later created and used a username. A significant part of the respondents, 38%, indicated that they preferred not to create a login account, although they knew they had the option to do so. Only 5 % of the respondents was not aware of the possibility to create a login account (see figure 13).

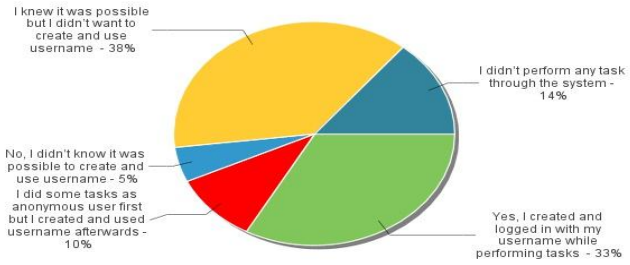


Figure 13, Questionnaire respondents' login use.

6. DISCUSSION

The results of this research project support our expectation that Nichesourcing is suitable for the digitization of pluvial data from Africa, and that people with African background are an appropriate niche group to be targeted for this purpose. This is indicated with recurrent participation of highly motivated niche as shown in the result section above (section 5).

The experiment was however implemented on a small scale with few atomic tasks for digitization, a short release period and very limited dissemination of the invitation for participation. At the beginning of this project, we anticipated to acquire more pluvial data from the region than the samples we have now used in the project. Due to the outbreak of a war in Mali, communication with regional contacts was hampered and more pluvial records could not be acquired. Based on the small number of available digitization tasks and on time constrains, we decided to run the system for the short period of one week and to invite a limited number of participants. The original idea was to announce the application on different online platforms of amongst others African Diaspora and African Youth, but instead an open invitation for participation was forwarded by email and through Facebook to and through the personal and professional contacts of the researchers.

Since logging in was not mandatory for participation some data related to anonymous users (not logged) digitization could not be collected. For instance the exact number of anonymous users and their recurrence in the digitization could not be computed. Moreover, we can only infer, but not totally proof that the respondents of the survey are the same people who participated in the digitization.

The assessment whether people with Africa related background are a suitable niche for digitization of pluvial data from Africa was performed to asses if we match tasks with the most appropriate expert. However, it was limited in two ways. First of all, the term 'people with Africa related background' is too general and not well defined. Second, the open invitation of participation was disseminated to and through personal contacts of the researcher. It is very well likely that people decided to participate because they personally knew the researchers.

This experiment is further limited because it mainly evaluated the reputed participation of users.

Future work

In order to establish whether Nichesourcing is suitable for the digitization of pluvial data from Africa and whether people with an African background are a suitable niche group for this type of project, further research will need to be conducted and implemented on a larger scale. Based on our experience we recommend the following for future research:

- ➢ Make the creation of a username and logging in mandatory for participation
- ➢ Investigate ways to make the entry of new data more efficient
- ➢ Evaluate the quality of the performed digitization
- ➢ Incorporate expert finding techniques [20]
- ➢ Include Gamification techniques that could complement quality of the result

7. CONCLUSION

In this research, we harnessed the power of a motivated crowd to digitize pluvial data from the Sahel region in Africa through Nichesourcing.

Despite the limitations discussed in the section above, we have produced thousands of structured digital data from handwritten pluvial records obtained from Mali in the Sahel region. We have also identified the prior anticipated niche, people with Africa related background, who actively participated in the project. From the participants who indicated to have an Africa-related background 'non Africans who have been living in Africa' were also active participants in addition to 'Africans who are living in Africa' and the 'African Diaspora'. We also found indications that a desire 'to participate in projects benefiting Africa' and 'to do good' motivates people more than 'earning points' or 'enjoyment'.

However these results might be biased due to the small scale of this project and the researcher's affiliation with Africa. Thus, we believe that further research is necessary to investigate whether the results also hold true when tested on a bigger scale (with more data and participants) to claim that Nichesourcing indeed suits such projects.

8. REFERENCES

[1] All Africa. Africa: Social Media Marketing in Continent. Available from http://allafrica.com/stories/201204190454.html (2012); accessed 29 August 2012.

[2] Balog, K., Azzopardi, L., and de Rijke, M. , A language modeling framework for expert finding. (2009)

[3] de Boer, V. , De Leenheer, Bon, A. , Gyan, N. , van Aart, C. , Gueret, C. , . . . Akkermans, H., . RadioMarché : Distributed Voice- and Web-interfaced Market Information System under Rural Conditions. (2012)

[4] de Boer, V. , Hildebrand, M. , Aroyo, L. , De Leenheer, P. , Dijkshoorn, C. , Tesfa, B. , and Schreiber, G. Nichesourcing: Harnessing the Power of Crowds of Experts. (2012)

[5] Foresman, C. All that user-generated content? 95% is malware, spam. Available from http://arstechnica.com/business/2010/02/all-that-use-all-that-user-generated-content-95-is-malware-spam-r-generated-content-95-of-is-malware-spam/ (2010); accessed 29 August 2012

[6] Heath, T., Hepp, M., and Bizer, C. . Linked Data - The Story So Far . Available from http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf (2011); accessed 29 August 2012

[7] Heer, J. and Bostock, M. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design (2010)

[8] Hoffmann, L. Crowd Control. In Communications of the ACM, Bosten, 2009 pp.17

[9] Liu, Y. , Alexandrova, T. , and Nakajima, T. Gamifying Intelligent Environments. (2011)

[10] Miller, J. , Bangor, A. , and Kortum, P. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies (2009) , 4(3). 114-123

[11] Miniwatts Marketing group,. Internet World Stats. Available from http://www.internetworldstats.com/stats.htm (2012); accessed 29 August 2012.

[12] Oomen, J. , and Aroyo, L. Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges (2011)

[13] Quinn, A., and Bederson, B. , . Human Computation: A Survey and Taxonomy of a Growing Field (2011)

[14] Sauro, J. , Measuring Usability with the System Usability Scale (SUS) (2011)

[15] Tesfa, B. , Regreening Africa App. Available from http://capp4ra2.ict2green.com/home (2012); accessed 29 August 2012

[16] Tesfa, B. Regreening Africa App on Facebook. Available from http://apps.facebook.com/regreening/ (2012); accessed 29 August 2012

[17] The World Bank. African Diaspora. Available from http://siteresources.worldbank.org/INTDIASPORA/Resources/AFR_Diaspora_FAQ.pdf (2012); accessed 29 August 2012

[18] Tomasic, A. , Zimmerman, J. , Steinfeld, A. , Huang, Y. ,Yoo, D. , Hiruncharoenvate, C. , and Ayoob, E. Design Uncertainty in Crowd-Sourcing systems (2011)

[19] Web Alliance for Regreening in Africa, About W4RA. Available from http://www.w4ra.org/en/about (2012); accessed 29 August 2012

[20] Wenger, E., McDermott, R., and Snyder, W. Cultivating communities of practice: a guide to managing knowledge (2002)

[21] Zooniverse. The power of the crowd. Available from http://www.oldweather.org/why_scientists_need_you (2012); accessed 29 August 2012

9. APPENDIX

Appendix 1: Invitation for participation

Are you willing to give a few minutes of your time to re-greening Africa?

Access to previous year's rainfall and harvest records is important for people working in the environmental and agricultural sectors. For example, farmers could better anticipate the coming year's rainfall based on the past year's trend and plan their harvest accordingly. Policy makers could also use harvest reviews of the past years to make plans for future years. However a lot of such records in Africa are only stored in hand written documents and can therefore not be accessed, reviewed and disseminated easily.

As part of my master's research I have developed an application through which you, together with other people from all over the world, can assist in digitizing documents from Africa. It only asks a few minutes of your time to copy hand written figures/text in to a digital form. In this way you will digitize a very small part of a hand written document. Your contribution will be combined with other people's contributions. When many people participate repeatedly, hand written documents can be converted in to digital format with high accuracy and within a short time.

I would like to invite you to participate in this release of the application which will run for about a week. The application can be easily accessed through Facebook Regreening App at http://apps.facebook.com/regreening/ or directly at http://capp4ra2.ict2green.com/ .

With your help we aim to digitize hand written rainfall and harvest records, from Sahel region in Africa, and make it available to support re-greening initiatives.

We appreciate it if you participate repeatedly within this week and also if you forward this message to people who might be interested in participating.

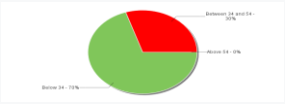
Thank you for supporting regreening Africa,

Appendix 2: Closing Survey Summary Report

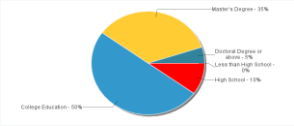
Summary Report: Regreening Africa App Closing Survey Review (Completion rate: 100.0%, 20)

1) Are you Male or Female?

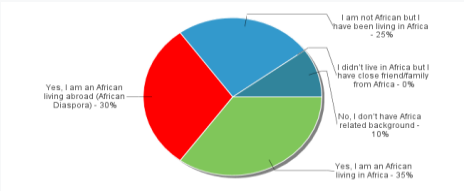
2) What is your age?



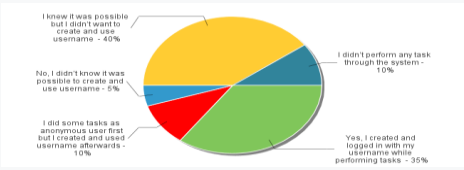
3) What is the highest level of education you have completed?



4) Do you have Africa related background?



5) If you had performed a task/s in the Regreening system, did you create a login account and login with your username/account (or use Facebook login) while performing the transcription?



7) Please indicate to what extent you agree with the following statements considering your motivation to participate.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total Responses	Mean	Median	Variance
7.1) I was motivated to participate in projects benefiting Africa	0 (0%)	0 (0%)	2 (10%)	4 (20%)	14 (70%)	20	4.6	5.0	0.5
7.2) I was motivated to participate in re-greening initiatives	1 (5%)	0 (0%)	2 (10%)	5 (25%)	12 (60%)	20	4.3	5.0	1.1
7.3) I was motivated to do anything good with the time	0 (0%)	1 (5%)	1 (5%)	7	11 (55%)	20	4.4	5.0	0.7

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total Responses	Mean	Median	Variance
I had				(35%)					
7.4) I was motivated because it was fun and enjoyable	0 (0%)	1 (5%)	7 (35%)	8 (40%)	4 (20%)	20	3.8	4.0	0.7
7.5) I was motivated because I received points and recognition	4 (20%)	3 (15%)	8 (40%)	4 (20%)	1 (5%)	20	2.8	3.0	1.4
7.6) I was motivated because I personally know part of the researchers and/or participants involved	1 (5%)	1 (5%)	2 (10%)	5 (25%)	11 (55%)	20	4.2	5.0	1.3

8) Please indicate to what extent you agree with the following statements considering the general usability of the system.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total Responses	Mean	Median	Variance
8.1) I think that I would like to use this system frequently.	0	0	7	9	4	20	3.9	4.0	0.6
8.2) I found the system unnecessarily complex.	5	8	5	2	0	20	2.2	2.0	0.9
8.3) I thought the system was easy to use.	0	1	3	10	6	20	4.0	4.0	0.7
8.4) I think that I would need the support of a technical person to be able to use this system.	9	6	5	0	0	20	1.8	2.0	0.7
8.5) I found the various functions in this system were well integrated.	1	1	5	10	3	20	3.6	4.0	1.0
8.6) I thought there was too much inconsistency in this system.	6	7	5	2	0	20	2.1	2.0	1.0
8.7) I would imagine that most people would learn to use this system very quickly.	0	0	3	11	6	20	4.2	4.0	0.4
8.8) I found the system very cumbersome to use.	2	9	6	3	0	20	2.5	2.0	0.8
8.9) I felt very confident using the system.	0	0	4	13	3	20	4.0	4.0	0.4
8.10) I needed to learn a lot of things before I could get going with this system.	7	6	3	3	1	20	2.2	2.0	1.6

Please let us know here if you have further remark.

Variable	Responses

1	Keep it up with the good work !
2	could be helpful to have a way to indicate whether a particular datum was illegible
3	Question 6; I thought I created a login account with my name and emailaddress. Later on however I realized that I should have used the password sent to me by e-mail. Therefore I would recommend to explain the login procedure more extensively. Furthermore I would recommend to create larger spaces for comments. Concerning the usability of the system I have to say that sometimes it was hard to read the handwritten text, so mistakes can easlily occur by digitizing those texts. Furthermore I noticed that the lay-out of the documents with the handwritten texts were not always exactly the same as the lay-out of the digital forms, which can create confusion (and again mistakes can occur). Overall I think the system will be very usable!
4	It wasn't clear that creating an account would be advanteagous for the reserachers

Appendix 3: Sample result: Rainfall record digitization for station bankass

Date and time	Digitizer	Document ID	Year	Title	Station	Month	ht/nj	Rainfall Value
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233,6
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233,6
01/08/2012 03:26	getanegash	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233,6
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233,6
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233.6
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233.6
25/07/2012 14:07	anonymous	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	ht	233.6
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
25/07/2012 14:07	anonymous	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
01/08/2012 03:26	getanegash	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13

30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Aoug.	nj	13
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206,9
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206,9
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206,9
01/08/2012 03:26	getanega sh	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206,9
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206.9
25/07/2012 14:07	anonymo us	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206.9
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juill.	ht	206.9
25/07/2012 14:07	anonymo us	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
01/08/2012 03:26	getanega sh	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juill.	nj	10
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81,3
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81,3
01/08/2012 03:26	getanega sh	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81,3

28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81,3
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81.3
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81.3
25/07/2012 14:07	anonymous	tlrf2008pl	2008	pluviometrie	bankass	Juin	ht	81.3
25/07/2012 14:07	anonymous	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
01/08/2012 03:26	getanegash	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Juin	nj	5
01/08/2012 03:26	getanegash	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20,5
27/07/2012 03:46	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20,5
31/07/2012 10:21	Giulia Manole	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20,5
28/07/2012 09:01	Marlijn Lelieveld	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20,5
25/07/2012 14:07	anonymous	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20.5
30/07/2012 04:17	chimok	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20.5
30/07/2012 07:48	chimok	tlrf2008pl	2008	pluviometrie	bankass	Mai	ht	20.5