

# The CHiC Interactive Task (CHiCi) at CLEF2013

Elaine G. Toms and Mark M. Hall

The Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield,  
S1 4DP, UK  
{e.toms,m.mhall@sheffield.ac.uk}

**Abstract.** The interactive task in Cultural Heritage in CLEF 2013 used a standardised interactive protocol, information retrieval system and interface to observe a set of participants remotely via the web as well as in the lab access an English language collection from the Europeana Digital Library. Both user response and log data were collected from the 208 participants.

**Keywords:** cultural heritage, Europeana, interactive information retrieval, user behavior

## 1 Introduction

The intent of the **CHiCi** task was to collect a sufficiently large enough data set that represented user interactivity with the European Digital Library collection so as to:

- a) model user search/browse behaviour and observe user interaction with culture and heritage materials, and
- b) build a collection of user-centred data that might be augmented and used in the future for testing various types of hypotheses about the process, the context and the nature of the interactivity, and in particular, be useful for simulated IR.

Rather than taking the classical explicit task-based approach, we used one user task: one with implicit goal that reflects the exploratory nature of interaction by a non-expert information consumer with culture and heritage information objects. As such it was designed to encourage user interactivity and immersion in a culture and heritage environment while the system collected as much data from and about the user as our systems would permit.

The highly structured process which controlled the unstructured task enabled multiple research questions to be asked, and potentially with the data set, multiple additional questions to be asked in future, such as: what do people do when exposed to such an environment? How does the search process change over the course of that immersion? How do people interact with the images and their associated metadata? What can we learn from a user “session”? Can we model such a pattern given other types of repositories?

## 2 Teams

Four teams participated in the interactive task, which required each team to process 30 participants via the web and 10 in a fixed, observable, lab-based location. The language of operation was in English, and all protocols and systems were expressed only in that language. The participating teams and the number of participants per team in web-based versus in-lab are illustrated in Table 1.

**Table1. Numbers of Human Participants per Research Team**

	<b>Web</b>	<b>Lab</b>	<b>Total</b>
Humboldt Universität	18	8	26
Royal School of Library and Information Science	12	19	31
Stockholm University	9	0	9
University of Sheffield	117	20	137
Other	4	1	5
<b>Total</b>	160	48	208

## 3 Research Protocol, i.e., the Lab Task

For the ChiCi task, one common experimental system developed by the University of Sheffield [4] (which was informed by previous implementations [12, 1] and the initial synopsis of such experiment [8]), one set of content, and one interface [6] was deployed and used by all teams. The ‘task’ thus was a multi-part protocol that extracted multiple types of data from participants and observed participants virtually in their interactivity with the system. The protocol was implemented by SIRE and followed the pattern outlined in Fig. 1. This protocol was used by all teams. The data was collected, stored and re-distributed back to teams by the University of Sheffield.

### 3.1 IR System and Interface

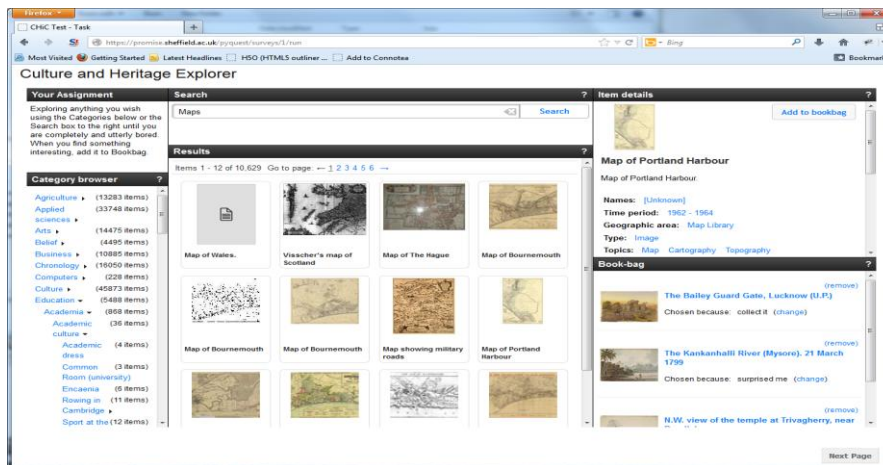
**Content:** The content contained 1,107,176 million records from the English-language collections of the Europeana Digital Library, as used in the other CHiC tasks.

**IR System:** The IR system was based on Apache Solr<sup>1</sup>, which provides the text search, spelling checker, and the “more like this” suggestions. The default settings were used for all components and all fields specified in the source records were loaded without any pre-processing.

---

<sup>1</sup> <http://lucene.apache.org/solr/>

In addition, a hierarchical category browser was added, based on the work of [2]. This process resulted in a set of 24 top-level categories, with between 3 and 14 sub-levels (median 5). The individual levels in the category hierarchy had between 1 and 384 sub-categories (median 3). A total of 267,768 items were automatically mapped into the category hierarchy. When the item – category mappings were loaded into Solr, each item was linked not only to the category the pre-processing had linked it to, but also to all of that category’s ancestors. When the user selected a category from the category browser, the Solr index was searched for all items that were mapped to this category. Because of the way the items were linked not only to their category, but to the category’s ancestors, this query would also return all items that were linked to the selected category’s descendants.



**Figure 1. Culture and Heritage Explorer**

**Interface – the Culture and Heritage Explorer:** Access to the IR system was provided using a novel interface (see Fig. 1); it offered three key ways of accessing the content: a) category browser, b) standard search box, and c) ‘more like this.’ The interface was a single display following a design used by [9].

The interface contained five key areas:

1) *Category hierarchy*: located on the extreme left, the hierarchy was navigated using the right arrow located to the right of each category, which expanded the level within the space, until it reached the nodes.

2) *Search box*: located in the middle, this was a conventional implementation of query entry that accepted keywords. After submitting a query, the results display was updated.

3) *Results display*: located below the search box, it displayed 12 thumbnails and titles of the thumbnails in a 3x4 grid layout. Navigation controls for moving from one results page to the other were provided above the display. When an item in the results display was clicked, the item appeared in the item display to the extreme upper right.

4) *Item display*: it contained the thumbnail and metadata fields associated with the item; unfortunately only the thumbnail is present in the data collection and thus no larger image was available.

The metadata associated with each item used the Europeana Data Model based on the Dublin core standard with Europeana-specific fields added. Some fields used expert jargon and were modified for a naïve participant (see table 2). At this point, an item could be added to the Bookbag using the button to the upper right. At the bottom of each item, the “more like this” displayed a variable number of thumbnail images linking to related items calculated using Solr's standard more-like-this functionality.

5) *Bookbag*: used for collected images that were deemed useful. Items in the Bookbag could be redisplayed or remove. Along with the item was the rationale for including the item as well.

**Table 2. Metadata fields displayed with each image**

<b>Metadata field</b>	<b>Label as Displayed</b>
dc:creator	Names
dc:type	Type
dc:subject	Topics
dcterms:temporal	Time period
dcterms:spatial	Geographic area
dcterms:isPartOf	Collection
Europeana:dataProvider	Object owner
Europeana:country	Location of object

*How the interface worked*: On startup, no query was inserted, but the results grid was populated with the first 12 documents in the Solr index to serve as a stimulus for starting the task. At that point, a participant could enter a query, scan the categories, select from those results. Once one item was displayed, a participant could also click on any metadata element to search by any of those metadata contents, select from more like this, or add an item to the Bookbag.

Once the “add to Bookbag” was selected, a popup box asked why the object was selected with the following options:

- I wanted to show someone
- I wanted to use the image in something
- I wanted to collect for a future purpose
- It surprised me!
- I simply liked it! No particular reason.

### 3.2 Experimental Task

The experimental task was presented to participants after they completed all preliminary questions. Prior to being assigned the task, participants were presented with a situation [2] to set the stage for the task:

“Imagine you are waiting to meet a friend in a coffee shop or pub or the airport or your office. While waiting, you come across this website and explore it looking at anything that you find interesting, or engaging, or relevant...”

The next page loaded the Explorer. The implicit task based on similar tasks used by [10, 11] and informed by [13] remained stationary in the upper left corner of the Explorer) was:

“Your Assignment: exploring anything you wish using the Categories below or the Search box to the right until you are completely and utterly bored. When you find something interesting, add it to the Bookbag.”

No further guidance was given, and participants were free to explore the resource; a mouse click on a ‘Next Page’ button disengaged the participant from the activity.

### 3.3 Protocol

Underpinning the protocol was a mixed methods methodology that required some pre and post the stimulus responses from participants, as well as the observation of the participants while doing the implicit task, i.e., the stimulus was activated. Thus the protocol extracted multiple types of data from participants and observed participants virtually in their interactivity with the system. The protocol followed the pattern outlined in Fig. 2, and was implemented as a sequence of web pages.

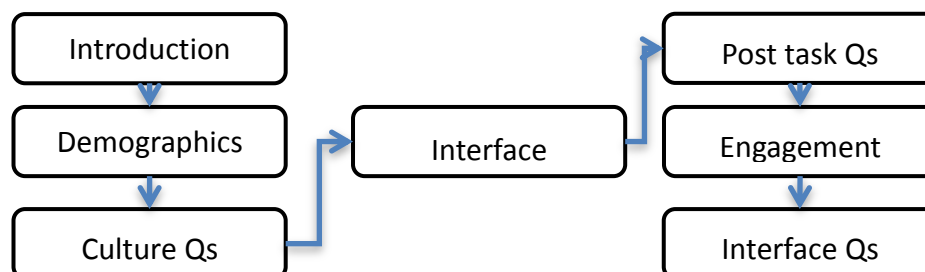


Figure 2. Sequence of Events in the Research Protocol

Participants accessed a remotely available system [4] that was tailored for this study. An information sheet and informed consent (required by the University of Sheffield's research ethics review process) was first presented to participants, followed by sets of questions about:

- basic demographic questions to create a profile of participant group;
- country of birth and residence, mother tongue, and language used to speak at home or search the web, to understand the potential impact of an individual's culture;
- museum visits, familiarity and interest in European culture and heritage and experience with the European Digital Library, to address whether the participant was 'of convenience' or interested in the topic matter.

While participants were engaged with the assigned experimental task, the system logged and time stamped the entire set of user actions and events including: queries, category selection, items examined, added to the Bookbag, and so on.

After the assigned task (see section 3.2), participants:

- responded to a 31 item User Engagement Scale [7] to assess the overall experience;
- provided a narrative explanation of why they included the objects in the bookbag, and their level of satisfaction with what they found;
- assessed the usefulness of each object on the interface;
- assessed the usefulness of each piece of metadata in assisting with assessing an item.

### 3.4 Participants

A synopsis of the participants is contained in Tables 3 and 4. The participant group contained a well-educated group of about 1/3 male. While the age range was wide, more than half were under 35, and thus a younger group. Participants came from 16 countries but more than half of those are resident in the UK, but originated, i.e., their birth, in 35 countries. Twenty languages are spoken today among the group, but they speak 26 languages in the home. Like country, language is dominated by English both as a mother tongue and as the current language spoken.

On a scale of 1 to 5 (from not familiar to very familiar), participants rated their familiarity with European culture and heritage at 2.2, and their level of interest European culture and heritage in the middle of the scale at 2.5. Of the participants, 78% indicated that they have never visited the European Digital Library (EDL) and 81% visit museums and galleries on the web or in person less than monthly.

**Table 3. Participants by Age and Gender**

<b>Age</b>	<b>Female</b>	<b>Male</b>	<b>Total</b>
18 - 25	53	24	77
26 - 35	47	24	71
36 - 45	16	12	28
46 - 55	12	8	20
56 - 65	8	4	12
Total	136	72	208

**Table 4. Participants by Age and Gender**

<b>Education</b>	<b>Female</b>	<b>Male</b>
	<i>IP/C</i>	<i>IP/C</i>
Secondary	0/97	0/53
Undergraduate	31/74	14/42
Further	1/79	4/40
Masters	33/39	14/23
Doctorate	14/14	17/11
Professional	1/7	1/2

Note: IP= in progress; C=completed

### 3.5 Procedure

People were recruited by each team using primarily email to participant remotely or to participate in a lab. The email link opened the system in a browser; Internet Explorer or Firefox were recommended, but the system was tested and worked in all major browsers. In addition to the protocol above, they were asked to indicate which team recruited them and whether they were currently in a lab or working remotely.

Those recruited to participate in a lab were processed independently as each team had an additional protocol. For example, the Sheffield team added a retrospective verbal protocol, which required participants to explain some of their decision-making. Other than this exception all teams used the same protocol and the same system and interface, and a central logger collected all log data.

### 3.6 Data Analysis

All user response data was collected to a relational database and exported as a CSV file for analysis. The task interface logged every interaction between the user and the system, including the output the system produced based on the user's request (see

sample format in Figure 2). For the analysis reported here, percentage, counts, means and SDs are primarily used. The more comprehensive analysis is contained in the individual lab papers, and in the follow-on analyses on the entire data set.

```
{
  "participant": 322, "timestamp": "2013-02-13T14:34:23",
  "action": "query", "parameters": {"q": "Railway"},
  "components": {
    "search_box": {"spelling": "Railway", "q": "Railway"},
    "search_results": {"numFound": 4, "docs": [{...}, {...]}}
  }
}
```

Figure 2. Sample log entry

Table 5. Use of interface objects.

		Web		Lab		Mean	
		#	SD	#	SD	#	SD
Queries	# of queries	3.5	8.6	5.3	6.6	3.9	8.2
Categories*	# of categories selected from hierarchy	9.3	11.3	19.6	22.8	11.7	15.3
Metadata facets*	# of metadata facets examined	0.7	2.1	2.4	6.4	1.1	3.6
Query Time	Time (sec) spent querying	187.5	60.4	234.3	253.1	198.1	541.2
Category time*	Time (sec) spent using categories	239.2	299.8	493.0	362.1	296.8	331.7
Metadata time*	Time (secs) spent using metadata facets	22.8	78.1	65.7	179.4	32.5	110.5
Interaction*	# of events/actions with system	57.1	63.4	97.1	67.6	66.2	66.4
Results page used*	# of results pages viewed	24.7	36.2	42.4	41.8	28.7	38.2
# = number of times, or time in seconds SD = standard deviation							



## 4 Results

From both user responses and the log files, we aggregated selected measures by participant. Because data were collected from two types of locations: via the Web and in the Lab, we present data by location as it became apparent in preliminary analyses that there maybe differences. But, because of the variation in size of the two location groups we are hesitant to say that these differences are statistically significant, and thus report the result and identify what looks suggestive (identified with an asterisk \*).

*How they interacted to find useful objects:* As illustrated in table 5, participants issued on average approximately four queries, examined almost 12 categories, and about one of the metadata items associated with each object. On average they clicked into something on the interface 66 times, and clicked through the results pages 28 times. Thus they interacted with all of the features, although some more than others.

*How they found selected objects:* They examined on average about 15 of the objects displayed, with about six of those resulting from queries to the system and seven emerging from using the category explorer. Of these objects approximately 6 (50%) were deemed interesting enough to add to the Bookbag (see Table 6). Of the objects placed in the Bookbag, on average 2.9 came via the category browser and 2.5 from querying the system.

		Web		Lab		Mean	
		#	SD	#	SD	#	SD
Bookbag	# of objects	6.0	8.3	4.5	4.2	5.7	7.6
Bookbag (category)	# of objects sved after category	2.9	4.7	2.5	3.1	2.8	4.4
Bookbag (metadata)	# of objects saved after metadata	0.3	1.3	0.3	0.7	0.3	1.2
Bookbag (query)	# of items in Bookbag after query	2.5	5.8	1.6	2.5	2.3	5.2

Table 6. Objects deemed useful to be added to Bookbag and source of these

*How they viewed the experience:* The were asked whether the objects were what would have been expected in such and repository and whether they were satisfied with these objects (see Table 7). Overall, they were dissatisfied with what they found, and found the objects they examined to be not what they would have expected of the EDL.

Table 7. Perception of and Satisfaction with the objects.

Measure	Definition	Web		Lab		Mean	
		#	SD	#	SD	#	SD
Expected	Scale of 1-5, degree to which objects were as expected	1.54	0.977	1.94	1.099	1.63	1.017
Satisfied	Scale of 1-5, degree to which objects were as expected	1.74	1.119	1.92	1.145	1.78	1.125

*Interface and Metadata:* In addition we wanted to understand the effect of the interface and the metadata associated with each object. We provided an image of the interface, identified each component, and asked about the usefulness of each of the objects in assisting with the exploration. Notably, like the question about satisfaction above, all were rated on the negative side on a five-point scale. Similarly, each object had a set of metadata associated with each, and of the set the Title, Description, and Thumbnail were considered to be somewhat useful in helping to assess the object with the title rated the highest at 2.8. Finally we asked an open-ended question: “What information would make the object in Europeana more interesting or useful to you?” Half indicated that the image as thumbnail was inadequate to view an object, and requested bigger, zoomable images so one could see the detail, and were clearly frustrated when they could not. Over a third commented on the lack of textual information associated with each object. Sometimes they wanted the context of where a picture was taken and in what year, and sometimes, more and richer descriptions of the object.

## 5 Discussion, Conclusion and Outlook

The results presented here are descriptive and summary. The interactive task was conducted in a highly controlled manner with a standard collection – the Europeana Digital Library, a standard IR System -- Apache Solr, and a standard interface – the Culture and Heritage Explorer, and used a standard research protocol. Data were collected from a wide range of countries/cultural groups both remotely on the web and in the lab.

The result of this process is a rich data set from user responses and from log file data. This collection of 208 responses can now be used to respond to a series of research questions that may relate to people and their interactivity, or to model that behavior and user action.

There are, however, challenges with how we interpret the data set:

*1) subtle differences between the remote and in-the-lab participants.*

With such a large difference in numbers: 48 in the lab, and 160 remote, it is impossible to know what caused the potential differences, or indeed whether those

differences are a spurious finding that would disappear in a randomly selected sample. Anecdotally, some in-the-lab participants volunteered that they stayed exploring because they were in-the-lab, but we stress, this is purely anecdotal. This requires follow-up.

Of all of the potential differences between use in the Lab versus on the Web, most notable was no difference in terms of interesting objects saved to the Bookbag. The differences appear, instead in the level of interactivity – both in aggregate and in use of the category explorer, suggesting being overseen in the lab may have changed their behavior, or doing the test off the web similarly gave them the anonymity to go through the motions of participating without commitment. The individual lab studies in which people came into the lab should illuminate this issue.

## *2) issues with the content*

Surprisingly, neither the interface nor the details associated with each object were considered useful in exploring the content. We wondered why, and the answer to this was clearly made by participants in their narrative responses to the question about what might be useful.

As described earlier, participants did not find the thumbnails useful and were under impressed with the limited amount of textual description. It seemed that the objects and their limited description asked more questions than answered them for participants. They commented and evaluated the category browser negatively, but without knowing that the category browser was relying on that very limited metadata detail to provide it with sufficient information to create it in the first place. We believe that the very limited content – a thumbnail, which in other repositories/databases is considered a preliminary view, and the extremely limited text when participants expected at least what one might find in a museum guide or in Wikipedia, negatively affected their overall view of the system.

As a result, we find this particular data set not very useful as a research resource for interactive information retrieval. It may serve other uses for automatic extraction, filtering, linking, etc., but in studies that require people to voluntarily interact with a resource, the resource has to have some interest for the participant.

## **Acknowledgements.**

This work was supported by PROMISE (Participative Research Laboratory for Multimedia and Multilingual Information Systems Evaluation, Network of Excellence co-funded by the 7th Framework Program of the European Commission, grant agreement no. 258191. We would like to thank Vivien Petras, Humboldt Universität, Nicola Ferro, University of Padova; Preben Hansen, University of Stockholm; Birger Larsen, University of Copenhagen, and Robert Villa and Paul Clough, University of Sheffield for their comments on various versions of the system and task. Last but not least, we would like to thank all participants (either in the lab or online) who generously gave their time.

## References

1. Bierig, R., J. Gwizdka, and M. Cole. A user centered experiment and logging frame- work for interactive information retrieval. In Proceedings of the SIGIR 2009 Work- shop on Understanding the User: Logging and interpreting user interactions in information search and retrieval, pages 8–11, 2009.
2. Borlund, P. The IRevaluation model: A framework for evaluation of interactive information retrieval systems. *Information Research*, 8(3). <http://informationr.net/ir/8-3/paper152.html>. (2003)
3. Hall, M.M., Clough Paul D., Samuel Fernando, Mark Stevenson, Aitor Soroa, and Eneko Aguirre. Automatic generation of hierarchies for exploring digital library collections. forthcoming.
4. Hall, M. & Toms, E.G. Building a common framework for IIR evaluation. *Proceedings of CLEF 2013* (in press).
5. Hall, M., Villa, R., Rutter, S., Bell, D., Clough, P., Toms, E. Sheffield Submission to the CHiC Interactive Task: Exploring Digital Cultural Heritage. In *Proceedings CLEF 2013, Working Notes* (2013).
6. Hall, M.M., Katsaris, S. and Toms, E.G.. A Pluggable Work-bench for Creating Interactive IR Interfaces. In *Proceedings of EuroHCIR2013, 2013*
7. O'Brien, H.L. and E. G. Toms. The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology*, 61(1): 50–69, 2009.
8. Tague-Sutcliffe, J. The pragmatics of information retrieval experimentation, revis- ited. *Information Processing & Management*, 28(4):467–490, 1992.
9. Toms, E.G., McCay-Peet, L., and Mackenzie, R.T. (2009). wikiSearch: from access to use. In *Proceedings of the 13th European Conference on Research and Advanced Technology For Digital Libraries (Corfu, Greece, September 27 - October 02, 2009)*. M. Agosti, J. Borbinha, S. Kapidakis, C. Papatheodorou, and G. Tsakonas, (Eds.) *Lecture Notes In Computer Science*. Springer-Verlag, 27-38.
10. Toms, E.G. (2000). Understanding and facilitating the browsing of electronic text. *International Journal of Human Computer Studies*, 52(3), 423-452.
11. Toms, E.G. (2012). *The relationship between curiosity and browsing patterns*. Presented to the Department of Computer Science, University of Leeds.
12. Toms, E.G., Freund, L. and C. Li. Wiire: the web interactive information re- trieval experimentation system prototype. *Information Processing & Management*, 40(4):655–675, 2004.
13. Wildemuth, B. & Freund, L. Assigning search tasks designed to elicit exploratory search behaviors. In Proceedings of HCIR 2012, October 4-5, Cambridge, MA, (2012). ([http://ils.unc.edu/searchtasks/publication/publication\\_2.pdf](http://ils.unc.edu/searchtasks/publication/publication_2.pdf))