

## New Methods in History

# A Life in the Trenches? The Use of *Operation War Diary* and Crowdsourcing Methods to Provide an Understanding of the British Army's Day-to-Day Life on the Western Front

RICHARD S. GRAYSON

*Goldsmiths, University of London*

Email: [r.grayson@gold.ac.uk](mailto:r.grayson@gold.ac.uk)

### ABSTRACT

*Operation War Diary, launched in 2014, is a crowdsourcing project in which 'Citizen Historians' tag First World War British Army Western Front war diaries to create data for analysis. This article discusses the methodology used in the project and then analyses (for the duration of the war) data for the six original British Expeditionary Force infantry divisions (1-6) and the first two cavalry divisions (1-2) to arrive in France. It highlights uses of the data on issues such as how much time was spent at the front and/or fighting, which appears to be 47% for infantry battalions, 62% for the artillery and 20% for the cavalry. At 46%, artillery days involving some fighting were more than twice the infantry's at 20%. The article also highlights further research questions and ways in which fuller datasets could be compiled using crowdsourcing methods.*

### **Introduction: Public and Academic Approaches to Day-to-day Life**

The BBC Schools World War One web page 'Life in the Trenches' is a useful indicator of what the public is most commonly told about fighting on the Western Front in 1914-18. Its opening line says: 'On the Western Front, the war was fought in trenches.'<sup>1</sup> There is no sense here of the mobile warfare of 1914 and 1918 and the role of the artillery located far behind the front line trenches, the biggest killers on the Western Front, is entirely absent. Although there is mention of the air war on a separate page there is no link between it and the fighting on the Western Front, nor is there any hint that certain types of soldiers, such as those in the Army Service

---

<sup>1</sup> <http://www.bbc.co.uk/schools/0/ww1/25626530> [accessed 1 May 2015].

Corps or the Royal Engineers, had very different roles to those of the infantry.<sup>2</sup> Crucially, there is no recognition that the infantry did not spend all of their time in the trenches.

The BBC Schools pages are merely the tip of an iceberg when it comes to popular representations of soldiers spending all of their days and nights in the trenches. The best known is, of course, *Blackadder Goes Forth*, which largely takes place in a trench dugout with only relatively brief forays to other places (primarily a staff headquarters, but also a hospital, a concert party venue, two jails, and a short period in the skies as members of the Royal Flying Corps). Throughout, 'home' for Blackadder and his men is a dugout, said by Private Baldrick to have been their location since at least New Year's Eve 1915. As the series is set in 1917 they appear to have been there at least for all that year thus far as well as all of 1916.<sup>3</sup> Quite how far the public believes that soldiers spent all their time in the trenches is something we just do not know. The only thorough study of public opinion on the war, a YouGov poll for the think tank *British Future*, unfortunately did not probe the issue. However, among a populace of whom the study showed one-third cannot name 1914 as the year the war began,<sup>4</sup> we might not expect very detailed knowledge of the nuances of day-to-day life beyond what Dan Todman calls 'the unremitting nature of trench warfare'.<sup>5</sup>

Aside from its schools pages the BBC has offered a different version of daily life in a piece fronted by the broadcaster Dan Snow and called 'How did so many soldiers survive the trenches?' This is itself a challenge to public perception of the war since its overall argument is that 'nearly 9 out of every 10 soldiers in the British Army who went into the trenches survived.' That statement is problematic because the commonly cited survival rate of around 88% applies across the Army as a whole, and includes servicemen who spent little or no time in the 'trenches'.<sup>6</sup> However, in making this case it usefully highlights the fact that not all of their time was spent in the trenches. Its (unsourced) figures for where life was lived by soldiers are: firing line (15%), support trench (10%), reserve trench (30%) and 'out of trenches' (45%).<sup>7</sup>

---

<sup>2</sup> <http://www.bbc.co.uk/schools/0/ww1/25401264> [accessed 24 July 2015].

<sup>3</sup> Richard Curtis, Rowan Atkinson and Ben Elton, *Blackadder: The Whole Damn Dynasty, 1485-1917* (London: Michael Joseph, 1998), pp. 441 and 451.

<sup>4</sup> *British Future, Do Mention the War* (London: British Future, 2013), p. 32, available online at: [http://www.britishfuture.org/wp-content/uploads/2013/08/BRF\\_Declaration-of-war-report\\_P2\\_Web-1.pdf](http://www.britishfuture.org/wp-content/uploads/2013/08/BRF_Declaration-of-war-report_P2_Web-1.pdf) [accessed 1 May 2015].

Dan Todman, *The Great War: Myth and Memory* (London: Continuum, 2005), p. 8.

<sup>6</sup> Jay M. Winter, *The Great War and the British People*, 2<sup>nd</sup> edition (London: Palgrave, 2003), p. 73.

<sup>7</sup> <http://www.bbc.co.uk/guides/z3kgjxs> [accessed 1 May 2015].

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

Perhaps there is not a better understanding of how soldiers spent their time because the subject is not one to which historians have so far given much attention. The most thorough study of any British Empire formation is of the 1<sup>st</sup> Australian Division by Robert Stevenson.<sup>8</sup> This work identified fifty-three different activities recorded in unit war diaries, grouping them into ten types of activity, and then three categories: administration, training and operations. Any such categorisation inevitably has its problems, but Stevenson's exceptionally thorough work reveals interesting overall figures.<sup>9</sup> Of the 1683 days of the division's existence, 886 (52.6%) of its days were spent on operations. Administration took up 423 days (25.1%) and training accounted for 374 days (22.2%). Looking further at the operations data it is striking that on only 29 days were 'offensive operations' the predominant task, while 'defensive operations' took up 750 days.<sup>10</sup> This division saw action twice in Egypt, in Gallipoli and on the Western Front. At Gallipoli, operations took up 86.2% of the days (243 of 282), compared to 57.6% on the Western Front (552 of 959) and 22.3% (43 of 193) during its two deployments in Egypt.<sup>11</sup>

For the British army, general references are made in a number of studies. Peter Simkins cites divisional histories suggesting of infantry units at the front 'that, in the normal routine of trench warfare, approximately half the infantry strength of a division would be at rest or in reserve'. He points out that this meant that even when on a 'tour of duty in the forward area' an infantryman would spend 'something between a quarter and half of his time in rest billets' which would be three to five miles behind the line. He then adds that divisions would periodically be withdrawn

---

<sup>8</sup> Robert Stevenson, *To Win the Battle: the 1<sup>st</sup> Australian Division in the Great War, 1914-1918* (Cambridge: Cambridge University Press, 2013), pp. 8-9 & 220-9. I am grateful to Prof. Peter Simkins for pointing me towards this work.

<sup>9</sup> In looking at a division as a whole, differences at battalion level will be missed, though conversely there are benefits to a divisional study in terms of breadth that a battalion level study cannot bring. Stevenson opted to count each day only once 'based on what appeared to be the division's main focus.' The critique of this methodology is the danger of omitting activities that only take up, say, 5% of each day, but do take up 5% of every day over a long period, so are not then shown at all in summary figures. The defence of the method is that unless the precise amount of time for each activity is measured (if it is even known) then simply noting that an activity occurred on a particular day is in danger of overstating the importance of that activity. The crucial question then becomes 'What am I trying to establish?' and it is clear that for Stevenson's questions, he chose the right methodology.

<sup>10</sup> *Ibid.* pp. 225-7.

<sup>11</sup> *Ibid.* p. 229.

'some twenty miles farther to the rear for a longer rest of three weeks or more'.<sup>12</sup> Simkins, along with Gary Sheffield and Niall Ferguson, cites Charles Carrington whose memoirs (written initially under the pseudonym of Charles Edmonds) note that in 1916, he spent 120 days in reserve, 101 at the front or close to it, 73 days resting, and the remainder of the year in hospital, on leave, or travelling.<sup>13</sup> Ferguson further argues that 'Three-fifths of an infantryman's time was in fact spent in the rear, not the front lines'. He cites the case of the 7<sup>th</sup> Royal Sussex Regiment spending 42% of their time in 1915-18 at the front or in close support.<sup>14</sup> Stephen Bull offers a different view, saying that the 13<sup>th</sup> York and Lancaster Regiment spent 51 consecutive days in 1916 on the frontline on the Somme. He also says that the 2<sup>nd</sup> East Lancshires usually spent 16-24 days in trenches at a time. However, he notes that 'from a few days to a week or so in the front line trench was a far more typical average'.<sup>15</sup> J.G. Fuller wrote of 'roughly three-fifths of the infantryman's service overseas spent in the rear of these lines [front line trenches]' and specifically described this as 'time spent in billets and rest camps'.<sup>16</sup>

A systematic sample has been carried out by Gordon Corrigan, who looked at five battalions for each January of the war over 1915-1918.<sup>17</sup> In all four months, he finds that no part of any battalion ever spent more than thirteen days in one month in the trenches and the longest continuous period spent in the trenches was six days with the longest continuous period in the firing line being four days. In all cases, these maximum figures occurred in January 1915. Meanwhile, although Corrigan did not work out any averages in presenting his figures, analysing his data further shows that the average number of days spent in trenches across all four Januarys was less than six days. If the exceptional January 1915 is removed, the figure falls to less than five.

---

<sup>12</sup> Peter Simkins, 'Soldiers and Civilians: Billeting in Britain and France', in Ian F.W. Beckett and Keith Simpson (eds), *A Nation in Arms: The British Army in the First World War* (Barnsley: Pen & Sword, 1985), pp. 165-192 [p. 178].

<sup>13</sup> Ibid. p. 179; Gary Sheffield, *Forgotten Victory, The First World War: Myths and Realities* (London: Headline, 2001), p. 150; Niall Ferguson, *The Pity of War, 1914-1918* (London: Penguin, 1998), p. 352. See also, Charles Edmonds, *A Subaltern's War* (London: Peter Davies, 1930), p. 120. Edmonds called the 101 days as 'days which may be described as under fire.' He went on to say 'This must be a typical experience shared by many hundreds of thousands of infantrymen who spent a year continuously at the front during the middle period of the war.'

<sup>14</sup> Ferguson, *Pity of War*, p. 352.

<sup>15</sup> Stephen Bull, *Trench: A History of Trench Warfare on the Western Front* (Oxford: Osprey, 2010), p. 94.

<sup>16</sup> J.G. Fuller, *Troop Morale and Popular Culture in the British and Dominion Armies 1914-1918* (Oxford: Clarendon Press, 1990), pp. 6 & 58.

<sup>17</sup> Gordon Corrigan, *Mud, Blood and Poppycock* (London: Cassell, 2003), pp. 90-93.

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

Other aspects of the trench experience have been addressed by Tony Ashworth and Mark Connelly. Ashworth has examined tours of duty in the trenches, focusing on whether tours were 'active' or 'quiet', finding that at least 40% were 'quiet' for the three battalions researched.<sup>18</sup> Both Ashworth and Connelly have also quantified raiding which was, of course, a relatively unusual activity.<sup>19</sup> Further work by Peter Simkins has addressed how many attacks took place on the Somme in 1916 by New Army Divisions, and also for the Hundred Days of 1918. However, that work asks different questions to those pertaining to daily life.<sup>20</sup>

### **Crowdsourcing and Methodology**

One good reason for historians not having systematically analysed the day-to-day lives of soldiers is that to obtain a meaningful figure for a range of units one would have to carry out a very laborious analysis over a wide range of unit war diaries with the purpose of examining daily activity. Nobody has sought to do so beyond Stevenson's analysis of the 1<sup>st</sup> Australian Division. However, the launch of Operation War Diary (OWD) by the National Archives, the Imperial War Museums (IWM) and Zooniverse in January 2014 has opened up a new opportunity to access and analyse precisely the broad range of data which is required.<sup>21</sup>

The project utilises digitised images from the National Archives' collection of British Army war diaries for the Western Front, totalling more than one million pages. War diaries are day-to-day records of the activities of units of the British army when deployed overseas. They are highly varied in terms of detail since they were written by unit adjutants with few formal guidelines over what they must contain. They tend to give fewer details on, for example, periods of resting or training compared to engagements with the enemy. Even the latter show great disparities in the amount of detail or comment included. Not every day is recorded in such a way as to allow them to be categorised in a consistent manner and days are occasionally missing, such as when a unit was overrun by the enemy and the diary lost. However, there is enough consistent terminology used in the diaries on the vast majority of days for categories of activity to be created. To do this OWD has adapted and extended

---

<sup>18</sup> Tony Ashworth, *Trench Warfare 1914-1918: The Live and Let Live System* (London: Macmillan, 1980), pp. 172-3.

<sup>19</sup> *Ibid.*, p. 185; Mark Connelly, *Steady the Buffs! A Regiment, a Region, & the Great War* (Oxford: Oxford University Press, 2006), pp. 246-50.

<sup>20</sup> Peter Simkins, *From the Somme to Victory: The British Army's Experience on the Western Front 1916-1918* (Barnsley: Praetorian Press, 2014), pp. 61-3.

<sup>21</sup> <http://www.operationwar diary.org/> [accessed 1 May 2015]; <http://www.1914.org/news/operation-war-diary-first-world-war-unit-diaries-go-online-to-mark-centenary/> [accessed 15 July 2015].

academic crowdsourcing techniques and software previously deployed by Zooniverse, principally in areas of science ranging from astronomy to climatology, although they have also been used in aspects of the humanities.<sup>22</sup> The core principle of crowdsourcing is that members of the public view evidence online and tag it in ways which make it accessible for analysis. The tasks and websites are designed to be accessible to anyone who can read English and has an internet connection.

Crowdsourcing techniques might be greeted with scepticism by historians accustomed to rather more solitary methods of research, in which 'expertise' measured by academic credentials is believed crucial to verifying facts. When historians use 'big data' it has usually been generated by official bodies (ranging from census returns to economic statistics), or has been created by other professional researchers. However, with funds for humanities research extremely limited at present the opportunities for funding teams of researchers to create data sets are now extremely scarce. Although not facing quite the same pressures, science has already recognised the value of volunteers and has shown that they can produce valuable and accurate data. Multiple volunteers making independent assessments might lead to errors but where there are errors they are not consistent enough to distort the results, and evidence from a range of projects suggests a strong tendency to come up with the correct answer, especially when set against the traditional 'gold standard'. One example of this is 'Snapshot Serengeti', a Zooniverse project launched in December 2012. That project examined data (in the form of photographs) created from 255 cameras positioned over 1,125 km<sup>2</sup> of the Serengeti National Park. Some 1.2 million sets of pictures were created and in analysing them more than 28,000 volunteers created 10.8 million classifications. When set against 'gold standard' classifications produced by experts, there was a 96.6% accuracy rate for species identifications produced by volunteers.<sup>23</sup> Many other projects can be found online<sup>24</sup> and there has been extensive rigorous analysis of why and how volunteers become involved in such projects<sup>25</sup> together with critical analysis of various initiatives.<sup>26</sup> The

---

<sup>22</sup> <https://www.zooniverse.org/#!/projects> [accessed 15 July 2015].

<sup>23</sup> Alexandra Swanson, et al., 'Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna', *Scientific Data* 2:150026 (2015). Available at: <http://www.nature.com/articles/sdata201526> [accessed 29 September 2015].

<sup>24</sup> <https://www.zooniverse.org/about/publications> [accessed 29 September 2015].

<sup>25</sup> See, for example, Lucy Fortson, et al., 'Galaxy Zoo: Morphological Classification and Citizen Science' in Michael J. Way, et al., eds, *Advances in Machine Learning and Data Mining for Astronomy* (Boca Raton, FL: CRC Press, 2012), pp. 213-36, available at: <http://arxiv.org/pdf/1104.5513v1.pdf> [accessed 29 September 2015]; Alexandra Eveleigh, et al., 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science', *CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 165

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

central partner in these projects is Zooniverse which is managed from the University of Oxford's astrophysics department. Zooniverse's methods are supported by a wide range of governmental and non-governmental bodies in the UK and the USA including the UK National Archives, the UK Met Office and the online digital and academic resource Jisc.<sup>27</sup> In addition to OWD, Zooniverse runs a further humanities project, Ancient Lives, which involves transcribing papyri from the remains of the Greco-Roman city of Oxyrhynchus in Egypt.<sup>28</sup>

The basic techniques developed by Zooniverse for verifying 'consensus' on data have been utilised for OWD and are discussed below. Meanwhile, the kind of 'expert' opinion which was applied to the Serengeti project through the creation of 'gold standard classifications' has also been provided for OWD. Working with the lead partners in the project, an academic group defined which types of activity should be tagged, and then a 'Field Guide' was produced and this is also discussed below.

For Operation War Diary the volunteers (labelled as 'Citizen Historians') were told that the data they gathered would be used for three key purposes:

- to enrich The UK National Archives' catalogue descriptions for the unit war diaries;
- to provide evidence about the experience of named individuals in the Imperial War Museum's 'Lives of the First World War' project; and
- to present academics with large amounts of accurate data to help them gain a better understanding of how the war was fought.<sup>29</sup>

---

(New York: ACM, 2014) pp. 2985-2994, available at: <http://discovery.ucl.ac.uk/1418573/1/p2985-eveleigh.pdf> [accessed 29 September 2015]; Alexandra Eveleigh, et al., "'I want to be a Captain! I want to be a Captain!': Gamification in the Old Weather Citizen Science Project", *Gamification '13: Proceedings of the First International Conference on Gameful Design, Research, and Applications* (New York: ACM, 2013), pp. 79-82, available at: <http://discovery.ucl.ac.uk/1412171/> [accessed 29 September 2015]; A useful blog discussion of a 2014 survey of 300 volunteers can be found at <http://blog.zooniverse.org/tag/study/> [accessed 1 October 2015].

<sup>26</sup> Mia Ridge, *Crowdsourcing our Cultural Heritage* (Farnham: Ashgate, 2014).

<sup>27</sup> <http://www.oldweather.org/about> [accessed 29 September 2015].

<sup>28</sup> <http://www.ancientlives.org/> [accessed 13 October 2015]. For one result see, '5156, Plutarch *Moralia* 660C, 661B-C (*Quaestiones Convivales* IV PR., 1.2)', in R.-L. Chang et al., eds, *The Oxyrhynchus Papyri*, LXXVII, (London, Egypt Exploration Society, 2012), pp. 97-98.

<sup>29</sup> <http://www.operationwardiary.org/#/about> [accessed 15 July 2015]. This link includes a full list of the Academic Advisory Group.

When accessing a war diary, volunteers can tag a wide range of items including date, place, type of activity, and names of individuals. From 1 January 2014 to 18 June 2015 (the cut off point for this study), there had been 612,962 classifications and these data have been posted online by the National Archives.<sup>30</sup> The data came from more than 12,600 registered users and more than 14,400 anonymous users giving over 27,000 classifiers in total. The most active user has made over 16,000 classifications, while the most active five users have made more than 10,000 classifications each. Around one-third of the users have submitted only one classification while just over half have submitted two or fewer. In keeping with established crowdsourcing standards, a war diary page is judged completely tagged once five users have examined it, and a war diary is 'complete' once all pages within it have been viewed and have been worked on by at least five people.

Differences are then resolved into 'consensus data' using a clustering algorithm with data on which there is no consensus being discarded. This not only allows errors in tagging to be removed but also increases the chances that all key information on a page will be tagged. The latter point is especially important because users have quite different approaches to tagging with some only focusing on place, person and date, while others are more comprehensive in their tagging of a wider range of activities on a daily basis.<sup>31</sup> An example of a complete page is included in Figure 1.

---

<sup>30</sup> <http://data.nationalarchives.gov.uk/wardiaries/> [accessed 2 October 2015].

<sup>31</sup> Information provided by Steven Hirschorn (The National Archives) in emails to author, 10 and 15 July 2015.



# OPERATION WAR DIARY AND CROWDSOURCING METHODS

Date	Location	Activity	Remarks
5 Sep 1914 (1)	Marsault	accommodation (3)	
5 Sep 1914 (1)	Marsault, Chateau (2)	0200PM (3)	Captain Preston (1)
5 Sep 1914 (1)	Chateau	movement (3)	
	Cherwi (5)	resting (1)	
	hot (2)		
0400AM (3)	Chateau (4)	0730AM (3)	
0300AM (3)	Serris (6)	1100AM (2)	0300PM (3)
	accommodation (5)	movement (2)	Lieutenant Hughes (6)
	movement (6)	resting (1)	Saint-Thibault (1)
	Crecy (6)	accommodation (1)	
	Vouilar	accommodation (1)	
	Maisonnelle (6)	hot (2)	
	Pierre Levee (6)	0700PM (3)	Lieutenant Wilson (6)
	Jouarre (6)	other (1)	reconnitered (1)
	La Ferté-sous-Jouarre	rain (7)	Lieutenant Wilson (1)
	Las Corbiers (6)	0700PM (3)	
8 Sep 1914 (1)	accommodation (3)	resting (1)	

Figure 1: 1<sup>st</sup> East Lancashire Regiment, 5<sup>th</sup> to 8<sup>th</sup> September 1914.  
Source: Operation War Diary.

## Tagging and Key Questions

This article is a start to the process of delivering on the pledge that OWD's data would 'present academics with large amounts of accurate data to help them gain a better understanding of how the First World War was fought'. It is written at the end of a significant phase of OWD but while the project is ongoing, partly so that some of the early results can inform debates taking place during the 1914-18 centenary period. The intention is also to show volunteers how their data is being used and to reflect on how tagging might need to develop in future to yield more comprehensive and accurate results.

At this article's core is data for the six British regular army infantry divisions and the two mounted divisions that were deployed to 'France and Flanders' in August and September 1914. OWD will, in time, produce much more data but the data currently available is roughly eight times as large as that which has been used for any other study, covering as it does eight divisions over the entire war compared to the largest previous study, Stevenson's work on one Australian division. Moreover, with the available OWD data relating to a discrete part of the British Expeditionary Force (BEF), a moment has been reached when data can be usefully assessed for a specific part of the army, the first to arrive in France, not least to inform debates which are taking place during the period of the war's centenary. The infantry data are most extensive and have been deemed worthy of focus for this reason as well: across all battalions there are 48,640 days (of a total of 60,479 tagged) on which

consensus was reached. This compares to 11,932 (of 15,413) for the cavalry, 5,990 (of 8,082) for the artillery, 5,327 (of 7,396) for the engineers and much smaller amounts for other types of unit. However, there is also a focus on the infantry because its data is most relevant to tackling public perceptions of 'life in the trenches.' At the 18 June 2015 date for generating data for these divisions, all infantry and cavalry data was complete, and that was also the case for all except a small selection of other units within each division. With the completion of data creation for the infantry and cavalry units of the first phase of the BEF a significant milestone has been reached. These data not only include the units which first went to France in August and September 1914, but any units that subsequently joined the division. So, for example, the 10<sup>th</sup> Gloucestershires only joined I Division in August 1915, but their diary from that point onwards is included.<sup>32</sup> It should be noted that war diaries only begin when a unit departs 'home' for deployment overseas, so figures are not affected by time spent in the UK prior to departure.

A vast range of data can be tagged, with date and place having been done especially thoroughly by the Citizen Historians. However, the classifications for 'unit activity' and 'army life' are crucial to determining what a unit was doing on a daily basis.

---

<sup>32</sup> Relevant orders of battle for the regular army divisions are at <http://www.longlongtrail.co.uk/army/order-of-battle-of-divisions/> and for the mounted divisions at <http://www.longlongtrail.co.uk/army/order-of-battle-of-divisions/order-of-battle-of-cavalry-and-mounted-divisions/> [both accessed 9 December 2015]. From the divisions examined, the units not completed are: I CAVALRY DIVISION: 1 Signal Squadron Royal Engineers; 9 Sanitary Section; 2 CAVALRY DIVISION: 4 Cavalry Field Ambulance; 7 Mobile Veterinary Section; 8 Mobile Veterinary Section; 9 Mobile Veterinary Section; Ammunition Park (56 Company A.S.C.); 'D' Battery Royal Horse Artillery; I DIVISION: 23 Field Company Royal Engineers; 26 Field Company Royal Engineers; 409 Field Company Royal Engineers; 1 Field Ambulance; 2 Field Ambulance; Divisional Train (7, 13, 16, 36, Companies A.S.C.); 2 DIVISION: Ammunition Column; 5 Field Company Royal Engineers; 226 Field Company Royal Engineers; 5 Field Ambulance; 6 Field Ambulance; Divisional Train (28, 31, 35, 172 Companies A.S.C.); 3 DIVISION: 11 Mobile Veterinary Section; 3 Divisional Train (15, 21, 22, 29 Companies A.S.C.); 8 Machine Gun Company; 4 DIVISION: 526 Field Company Royal Engineers; 12 Field Ambulance; 11 Field Ambulance; 3 'A' Sanitary Section; 4 Mobile Veterinary Section; Brigade Machine Gun Company; 5 DIVISION: Divisional Signal Company; 13 Field Ambulance; 6 Sanitary Section; 5 Mobile Veterinary Section; Divisional Train (4, 6, 33, 37 Companies A.S.C.); Brigade Machine Gun Company; 6 DIVISION: 2 Brigade Royal Field Artillery; Divisional Signal Company; 8 Sanitary Section; Brigade Machine Gun Company.

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

There are twenty-one different 'activity tags', defined by the OWD project team in conjunction with academic experts. For the purpose of analysis here, the author has grouped sixteen of them under three headings, with the other five as categories in their own right. This is partly so as to assess broad patterns, but also as a way of reducing the effect of different volunteers using different tags for the same type of activity. The groups are:

### **Fighting**

- *Achieved objective*
- *Attacking/firing*
- *Raiding*
- *Under fire*
- *Withdrawing*

### **Front line**

- *All quiet*
- *Digging/repairing positions*
- *Enemy activity*
- *In the line*
- *Patrolling & reconnoitring*

### **In reserve**

### **In support**

### **Movement**

### **Other**

- *Casualty treatment*
- *Clearing & burials*
- *Construction (not front line or gun positions)*
- *Other*
- *Re-supplying*
- *Working party*

### **Resting**

### **Training**

Definitions of each of the twenty-one tags have been provided for volunteers in an online 'Field Guide' that is central to the training of Citizen Historians.<sup>33</sup> Although we cannot be sure how accurately they have been followed, the grouping mentioned above will reduce the impact of some differences. Meanwhile, there has been frequent subsequent discussion between volunteers, and the provision of answers to

---

<sup>33</sup> <http://www.operationwardiary.org/#!/guide> [accessed 15 July 2015].

common questions, and this might mean that tagging has become more consistent across users as the project has developed.<sup>34</sup>

Some explanation of the three broad categories used here is necessary. 'Fighting' includes any moment at which there is some record of soldiers attacking or coming under fire. They need not have been in the front line when this happened but these moments are grouped together as the most vivid and dangerous aspects of war. The 'front line' category is used to gather any information on when soldiers were in front line trenches or otherwise (such as at times of non-trench mobile warfare) at a geographic point that can be considered as being the 'front'. 'Other' covers a wide range of tasks that would usually take place behind the very front of the line. Of the remaining categories, it was useful to keep 'in reserve' and 'in support' distinct from 'front line' because, although often trench-based, such times were a different type of activity to being in the immediate front. Meanwhile, 'movement', 'resting' and 'training' all involve such activities carried out behind the lines.

In addition, there were ten 'army life' tags, also detailed in the 'Field Guide':

- Accommodation
- Sport & Leisure
- Religion
- Rations
- Parades
- Inspections
- Discipline
- Medical
- Uniform
- Hygiene

The 'army life' tags have been applied by users relatively infrequently and in the data presented below, they have been grouped under the single heading of 'domestic'. Many of these tasks could be carried out while at the front. However, they are only counted as the activity for a day when *only* that activity took place, which is more

---

<sup>34</sup> See, for example:

<http://talk.operationwardiary.org/#!/boards/BWD000000g/discussions/DWD0000lts>,  
<http://talk.operationwardiary.org/#!/boards/BWD000000h/discussions/DWD000009z>  
and  
<http://talk.operationwardiary.org/#!/boards/BWD000000h/discussions/DWD0000jf7>  
[all accessed 2 October 2015].

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

likely to indicate that the men were behind the lines. Certainly this applies to parades which were the largest single category of 'domestic' activity.

The crucial questions for this study are around how often men fought and how much time they spent at the front. Therefore, any day on which men fought has been considered a fighting day (and only a fighting day, even though they would usually have rested at some point). Following this, any day at the front where there was not fighting has been given a separate category in the analysis. This does mean that where a unit was at the front in an evening and relieved overnight in the early hours of the next day, those two days will both be considered as days spent at the front. While specific temporal information is often captured for set piece fighting, the duration of other activities is rarely recorded, and it is therefore not possible to get an accurate measure of how much time was spent on each domestic activity. Meanwhile, if only part of a battalion was involved in an activity, for example fighting, then that is included as fighting for the whole battalion, partly because there is no way in OWD of splitting activity within a battalion, but also because it is not always clear how much of a battalion was specifically involved. However, both this and the point about 'overnight' relief just mean that one needs to be aware of the central question in this research which is 'according to unit war diaries, on how many days did X take place for a unit or part of it?' where X will give priority to any time fighting or being at the front.

Five more categories (reserve, support, resting, movement and training) are counted where men did only those activities, as with 'domestic' activities as explained above. This might at first be thought to be unusual because resting might take place on most days, and training might also take place as part of movement. However, the nature of the notes in war diaries, and the ways in which users have tagged them, means that there tends to be only one activity on most days. Finally, a category defined in the data analysis as 'residue' covers all those days where either an 'Other' task was done, or the day consisted of a range of activities (but not including fighting or being at the front line).

### **Data from OWD: Some First Results**

Any data analysis at this stage comes with caveats. Averaging out over six infantry divisions might well obscure substantial differences of experience between battalions. For that reason, some data is analysed later on 'raiding' to illustrate differences within and between divisions. However, if we are to understand thoroughly the full range of experiences then we need data for divisions beyond those in the first wave of the BEF. This will have to wait until much more work has been completed by the OWD volunteers. A further caveat is that there are sizable elements within some datasets where consensus has not been reached. The possible implications are

discussed below but numerous general patterns can already be seen in the current data.

Beginning with the infantry, the data covers 123 battalions for their time in 1-6 divisions, a vastly greater sample than has ever been used before, and a number of points emerge as seen in Figure 2.<sup>35</sup> Days on which fighting occurred were 20% across the entire war. The period with the highest proportion of days on which it took place was the quarter from October to December 1914, at 31%. Only in six other periods was it at 20% or more:

---

<sup>35</sup> Infantry battalions are not included for any period when formally designated as 'pioneers'.

# OPERATION WAR DIARY AND CROWDSOURCING METHODS

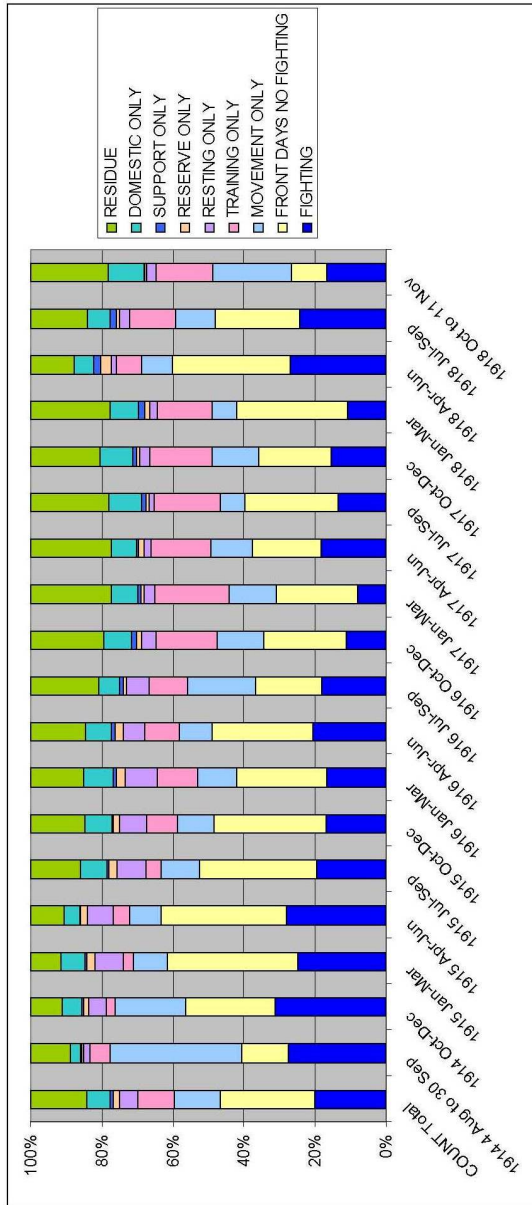


Figure 2: Infantry service, consensus diary. Source: Operation War Diary.

August to September 1914 (27%), January to March 1915 (25%), April to June 1915 (28%), April to June 1916 (21%), April to June 1918 (27%) and July to September 1918 (24%). For 1914, this reflects the six divisions' various involvements at the battles of Mons, Le Cateau, the Marne, Aisne and Messines. They were then at Second Ypres, Aubers, Festubert and Givenchy in the spring and early summer of 1915, and at the battles of the Second Somme and the Hindenburg Line in September 1918. The lowest incidence of fighting was January to March 1917 at 8%. Time at the front without fighting was highest in January to March 1915 at 37%. However, taking fighting and other time at the front together, the peak was April to June 1915 at 63%, followed by April to June 1918 at 60%. Across the war as a whole, the figure for being either at the front or fighting comes to 47%.

Looking at other areas of data, it is no surprise to see that, for the infantry, periods of movement were at their highest in August to September 1914 (37%) and October to November 1918 (22%), the periods when the front was extremely fluid with units being moved around to plug or exploit gaps in the line across a broad front. The result of this was that battalions would find themselves moving much longer distances than usual for the rest of the war without necessarily engaging the enemy. Other peak periods for movement were October to December 1914 at 20% and 19% in July to September 1916. Note again that 'movement' does not mean that troops were moving forwards or backwards or gaining ground due to attacks or defences (data for such days would be covered by fighting), but being moved from place to place, usually towards or away from deployment. All of these periods of movement coincide with intensive periods of fighting for the BEF as a whole.

Periods of infantry training are also notable for their differences. At the end of December 1915 it had never reached a double-digit percentage. From that point, it only failed to do so once (April to June 1918). 1917 was a highpoint for training with 21% in January to March, and an average of 19% for the year compared to 10% across the war as a whole. To some extent, this mirrors a decline in the days spent only resting. At 5% across the war, it was never below 5% from October 1914 to September 1916, but from then on resting ranged between 1% and 4%.

What of non-infantry units within divisions 1-6 and cavalry divisions 1 and 2? For the artillery (Figure 3), the range for fighting days across the war is higher than for the infantry, at 46% compared to 20%. The additional time spent in the artillery's 'front', which of course could be far behind the infantry's front, is only another 16%. This means that total days in an artillery 'front' position, whether fighting or not, is 62%. It is certainly plausible that such figures would be higher for the artillery than the infantry given the very different natures of their engagements with the enemy. The conditions of an artillery position compared to an infantry trench meant that much longer periods in the former were possible.



## OPERATION WAR DIARY AND CROWDSOURCING METHODS

In contrast to the artillery 'fighting' for longer than the infantry, we would expect the cavalry to be doing so far less often and that is what the data in Figure 4 shows. 'Cavalry' here includes not only the two cavalry divisions but also the mounted units of divisions 1-6.<sup>36</sup> Across the war as a whole, the data suggest that only 9% of the cavalry's days involved fighting. The most active fighting segments of the cavalry's war were August to September 1914 (27%) and October to December 1914 (25%). From the end of June 1915 it did not rise above 10% until January to March 1918 when it reached 15%. The final months of the war, October to November 1918 saw the cavalry fighting on 11% of their days. Meanwhile, time at the 'front' when not fighting was not nearly so significant for the cavalry as the infantry because of the needs of keeping horses. Across the war, it was just 14%, meaning that the total for days spent fighting or at the front was 23%. As for the infantry, cavalry training was much greater in the second half of the war. The largest category for the cavalry is 'residue' consisting of a range of behind the lines tasks at 22%.

The Royal Engineers (Figure 5) had a reverse pattern to the cavalry in that their time at the front was overwhelmingly dominated by periods in which they were not involved in any fighting.<sup>37</sup> Across the war, they spent 44% of their time at the front of which just 5% saw fighting. Peaks for time at the front came in July to September 1915 (65%) and April to June 1915 (60%), which can be seen as a period in which there was much consolidation of positions. Fighting for the engineers was much more constant throughout the war than for other units with no marked difference between mid- to late 1918 and other phases after 1914. For the Field Ambulance units (Figure 6), exposure to fighting was almost twice as much as for the Engineers at 9%. However, they spent far less time at the front when not fighting, just 5%, meaning that of all the BEF units examined so far, the Field Ambulance Units appear to be the least likely to have been at the front.

---

<sup>36</sup> B Squadron Northumberland Hussars in 1 Division (August to October 1916), A Squadron Northants Yeomanry in 4 Division (July to December 1916) and the Household Battalion also in 4 Division (November 1916 to February 1918).

<sup>37</sup> Field Companies, rather than Signal Corps.

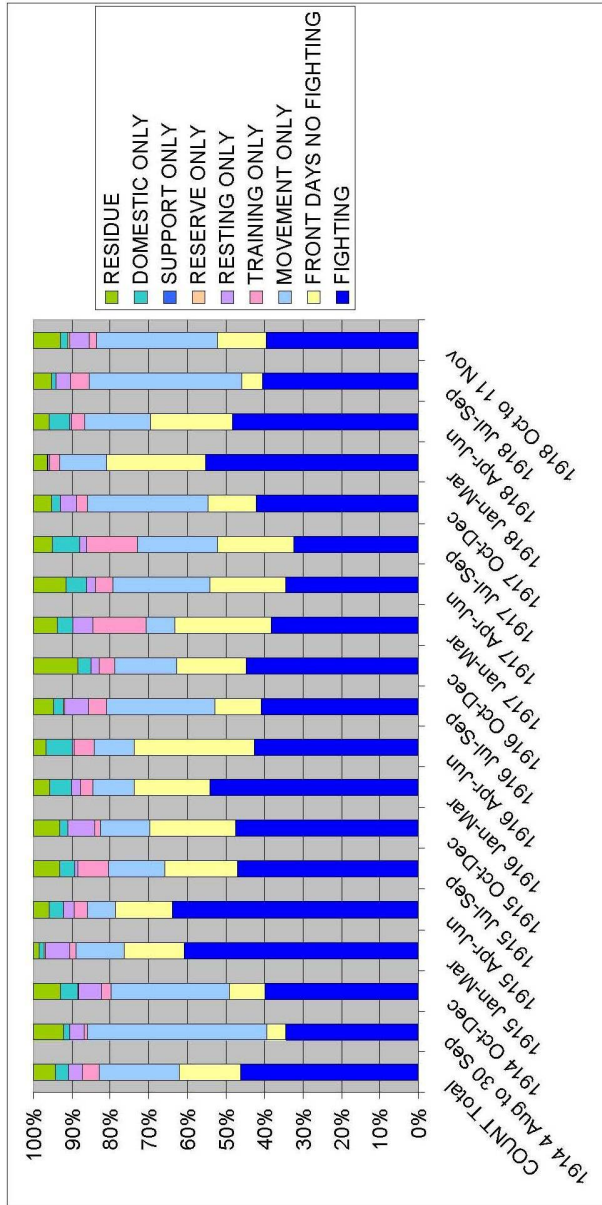


Figure 3: Artillery service, consensus data. Source: Operation War Diary.

# OPERATION WAR DIARY AND CROWDSOURCING METHODS

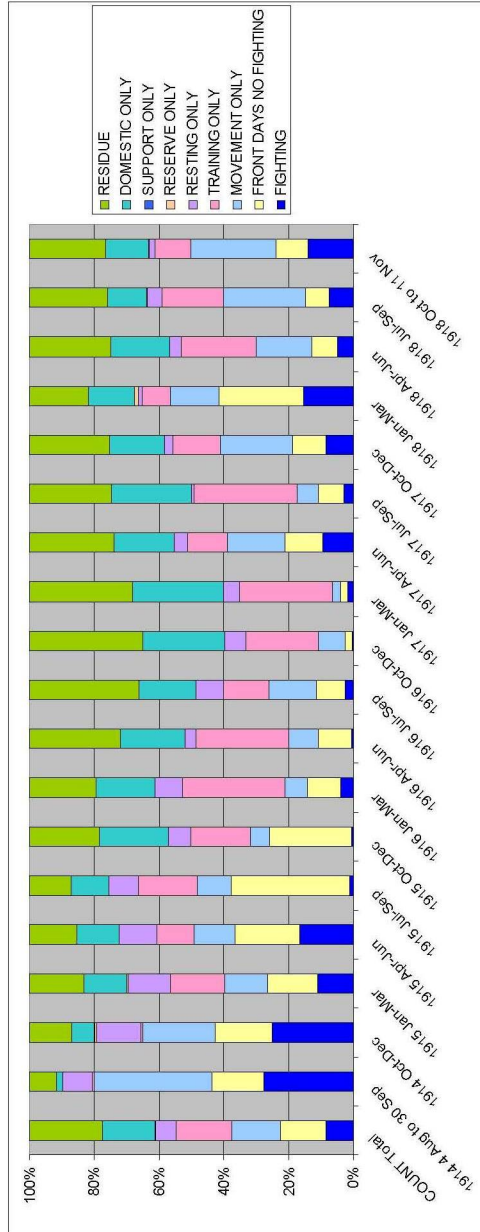


Figure 4: Cavalry service, consensus data. Source: Operation War Diary.

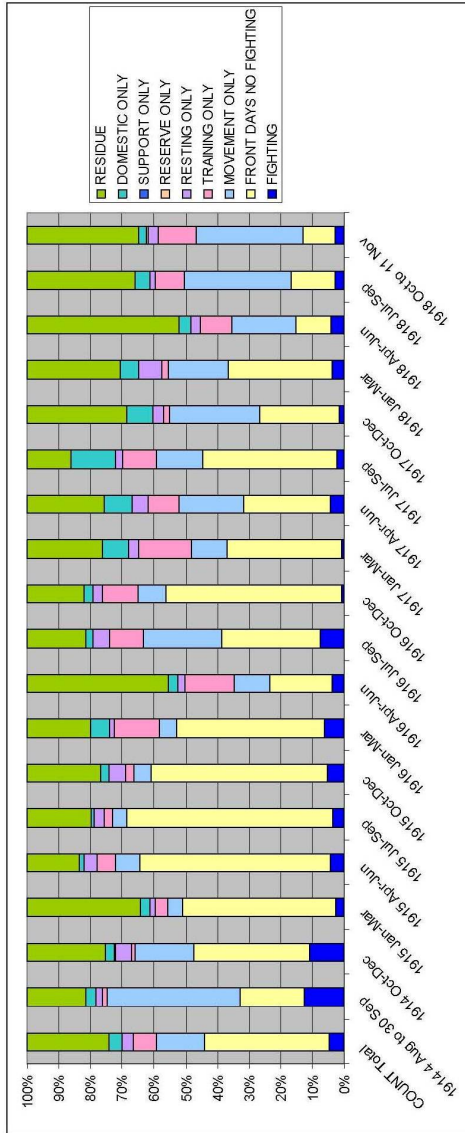


Figure 5: Royal Engineers service, consensus data. Source: Operation War Diary.

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

Within these figures there will of course be numerous exceptions to general patterns. Although battalions would have been at the 'front' for certain periods, the four companies of each battalion would be rotated. As explained earlier, the OWD data-mining interface does not extract information about differences in activity within battalions by company (indeed, it is not often very clear in battalion war diaries), and this is likely to be why figures for 'reserve' and 'support' seem to be far too low. However, long runs of particular activities for a battalion can be examined by following up OWD tags and looking closely at relevant war dairies where there seem to be unusually long periods at the front. Despite the overall figures, some units could have very long periods of duty there without any or at least lengthy relief. Among the battalions tagged in OWD, the 1<sup>st</sup> East Lancashires had a lengthy period almost entirely in the line, between 21 October 1914 and 11 April 1915. Having taken the village of Le Gheer near Ploegsteert on 21 October 1914, the battalion held it until 4 November (with A and B companies rotating with C and D in the very front line), by which point trenches had been established. After relief by another battalion on that day, they were in reserve trenches until 7 November, when they came under attack from a German attempt to retake the positions gained by the battalion a few weeks before. The 1<sup>st</sup> East Lancashires were then under shell fire for ten days, until they were relieved, going to Nieppe about three miles behind the lines. They were there for just three days before going back into the line on 20 November, where they stayed for almost exactly four months, with different parts of the battalion being relieved on 19 and 20 March. In all that time, there is one note of one company being relieved by a company of the London Rifle Brigade (on 27 December). It may well be that the diary did not include all such reliefs, but it does remain the case that at least part of the battalion was there for a third of a year, a period consisting of hard winter months. Incredibly, after just four days behind the lines, the battalion was back in at Le Gheer from 24 to 30 March 1915, before what might be considered a more usual period of rotation taking effect and a week or so at a time in the line at a range of different places.<sup>38</sup>

---

<sup>38</sup> The National Archives, WO 95/1498: 4 Division, 1 Battalion East Lancashire Regiment (1 August 1914 to 31 January 1918).

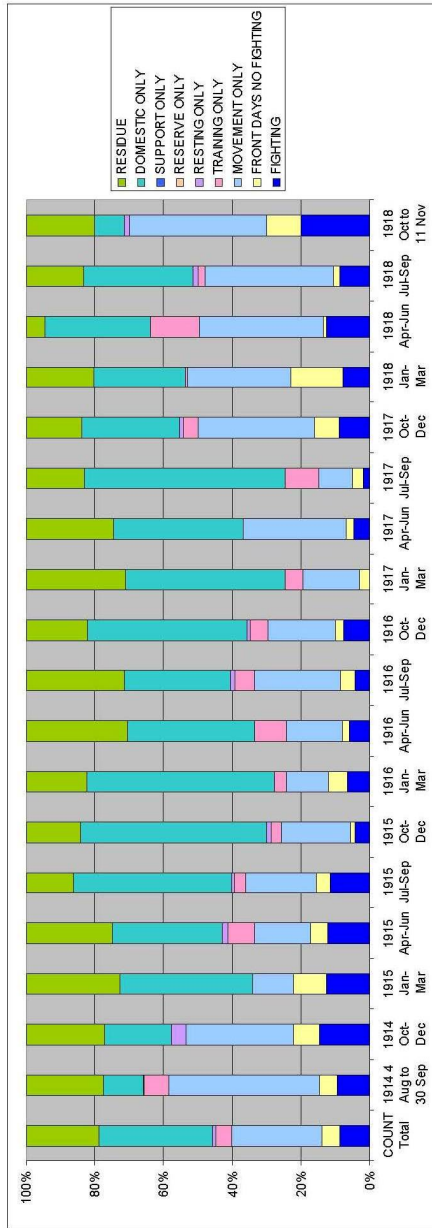


Figure 6: Field Ambulance, consensus data. Source: Operation War Diary.

## OPERATION WAR DIARY AND CROWDSOURCING METHODS

Deeper analysis of a range of battalions needs to be undertaken to answer questions including:

- To what extent and in what ways did infantry battalion experiences at the front vary from unit to unit?
- Are different patterns of service observable in different sectors?
- Was there change in patterns of service over time?

There are also questions to ask around how far certain activities can be taken to indicate that a division or battalion had some kind of elite status. One way of assessing this is to examine the occurrences of raids. By division (with battalions that carried out five or more raids specified), the following numbers of raids are recorded in OWD for the duration of the war:

### **1 Division (20)**

- 1<sup>st</sup> South Wales Borderers: 5

### **2 Division (33)**

- 1<sup>st</sup> Royal Berkshire Regiment: 8
- 13<sup>th</sup> Essex Regiment: 6

### **3 Division (30)**

- 2<sup>nd</sup> Suffolk Regiment: 6
- 8<sup>th</sup> King's Own Royal Lancaster Regiment: 6

### **4 Division (37)**

- 2<sup>nd</sup> Duke of Wellington's (West Riding Regiment): 8
- 2<sup>nd</sup> Essex Regiment: 12

### **5 Division (42)**

- 14<sup>th</sup> Royal Warwickshire Regiment: 8
- 1<sup>st</sup> East Surrey Regiment: 5
- 16<sup>th</sup> Royal Warwickshire Regiment: 5

### **6 Division (31)**

- 11<sup>th</sup> Essex Regiment: 6
- 2<sup>nd</sup> Durham Light Infantry: 7

From the figures above, it can be seen that 1 Division carried out around half as many raids as each of 4 and 5 Divisions. Within the divisions, the 2<sup>nd</sup> Essex Regiment led the way with twelve raids, and another three battalions with eight. However, as seen in Figure 7, in divisions 1-4, half or fewer of the battalions in each division took part in raids, with the rest apparently taking no part. In 5 Division, around two-thirds of the battalions did so and the British generals appear to have regarded this division as a good raiding force, even if the German assessment was that the division performed as merely average as attackers. Of divisions 1-6, none were judged in early 1918 to be 'Besonders gute Angriffsdivision' ('Particularly good assault

division'). However, divisions 1, 2 and 3 were seen as 'Gute Angriffsdivision' ('Good assault division'). Divisions 4 and 5 were regarded as average, while 6 Division was given the lowest ranking of 'Mittelmässige Division' ('Mediocre Division').<sup>39</sup>

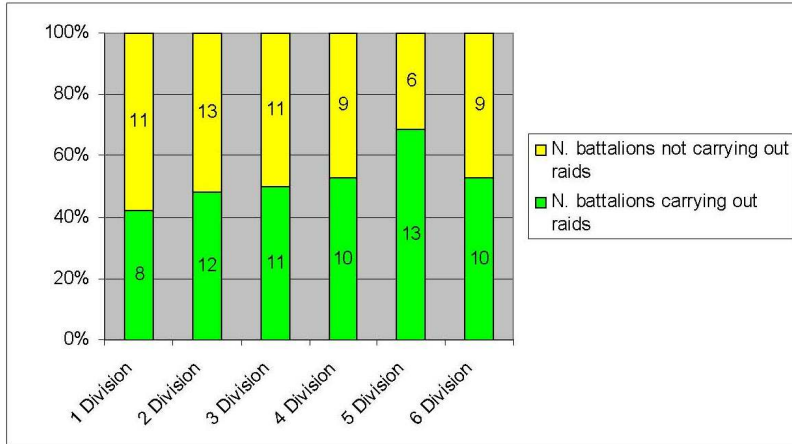


Figure 7: Incidences of battalions carrying out raids at least once, by division.

Source: Operation War Diary.

## Data Challenges

There are two obvious problems with the data which has emerged so far. In the first place, for the infantry, days spent in 'reserve' and 'support' positions are low at 1% each and fly in the face of accounts by, for example, Charles Carrington of how much time was spent on different activities. As stated above, a possible explanation for this is that it is not possible to tag different activities within a battalion for different companies, so the rotation of companies between the front and reserve lines would not show up. It is also possible that any time spent in the trenches (including reserve/support) has been tagged as in the line, but we would need to see data from volunteer divisions to get a clearer picture of this. These data might reflect the nature of the diaries themselves with the less stressful or less eventful activities being underreported.

Potentially a bigger problem, but one which remains by necessity unknown, is the lack of consensus on 20% of the days for the infantry, and in one quarter (July to September 1915), the figure is as high as 25%. It is a little higher for some other types

<sup>39</sup> Hauptstaatsarchiv Stuttgart M33/2 Bü 536: Mitteilung über die britische Armee Nr. 4, 1 January 1918. I am grateful to my Goldsmiths colleague, Professor Alex Watson, for providing the source and translations.



## OPERATION WAR DIARY AND CROWDSOURCING METHODS

of units (cavalry 23%, engineers 28%, artillery 26%), with the field ambulance units something of an outlier at 50%. What if all of these days were actually for one type of activity, or for a few types? To consider this, we need to be clear about how consensus is reached. In its simplest terms OWD deploys a majority vote system to derive 'consensus' from the data. A tag on which there was not consensus was either created by just a single user or was tagged by multiple users but without there being a majority for one activity over the others. So, for example, if four users tagged an activity, and two each took a different view, no consensus was reached but if five tagged an item, with three taking one view and two another, then consensus was reached. It is possible that the lack of consensus does apply to certain types of data more than to others, where the type of activity was somewhat ambiguous. However, that is unlikely to apply to more dramatic events such as fighting, so figures above for time spent fighting are likely to constitute a maximum.

One way round this would be to explore classifications having only one tag. Some of the single classifications are likely to have come from very detail-oriented individuals who were exceptionally thorough. Using such an approach, one would analyse how many single classifications there are and then rigorously spot-check those to see if they appear to be reasonable. The extent to which some tags were the subject of disagreement more than others could then be examined. One could also ask whether single tags were more likely to require some kind of interpretation beyond simply labelling a 'fact', and therefore were only contributed by the most engaged volunteers. However, this is a task for a later stage of the project.

### **Conclusion**

Operation War Diary has so far only produced a small proportion of the data that it will be able to offer to historians. However, the volume of the data is already so large as to go far beyond the amount used in previous studies and it is already clear that the data is of value in looking at overall patterns of BEF service. It can show differences over time for different types of unit, and it can show differences between such units. There is further scope for delving more deeply into these data sets to look at, for example: individual battalions or 'specialist' units of the BEF that have not been discussed here (such as Trench Mortar Batteries, machine gunners, cyclists, and veterinary sections). It would also be possible to narrow down the scope of the study to specific parts of the front.

Given historians' past analysis of the war, and evidence from memoirs, it may well be that reserve/support time has been seriously underestimated, not least because war diaries are often lacking in details when it comes to company level activity. However, until data is available for a much wider range of divisions, specifically the volunteer divisions, then we should allow for the possibility that the infantry from divisions 1 to 6 did spend far more time at the front than infantry from other divisions. It is

certainly possible that they did so. Paddy Griffith has pointed out that the five infantry divisions deployed in August 1914 were sent into 'serious action' on at least thirty occasions throughout the war. Many volunteer divisions saw only about half as many major actions, even though they were in France or Belgium for much more than half of the war. Thus there is some sense that at least some of the early divisions to arrive were seen as 'elite' although there was no precise definition of what that meant.<sup>40</sup>

Irrespective of any underestimate of reserve/support time or questions of the possible elite status of the divisions so far studied, the figure of 47% of infantry time being spent at the front or in fighting (with fighting occurring on 20% of those days) is a challenge to public perceptions. It also contrasts with the figures of 62% for the artillery and 23% for the cavalry. Furthermore, the artillery's 46% of days spent fighting compared to 20% for the infantry and 9% for the cavalry is a stark contrast, even if the nature of their fighting was different.

As the first example of a major crowdsourcing activity being applied to military history there are plenty of reasons to believe that Operation War Diary can help historians and the public to understand more about both the patterns and nuances of army life, and whether there was indeed a 'life in the trenches'.

### **Acknowledgements**

I am very grateful to a number of people involved in Operation War Diary (OWD), especially to Steve Hirschorn of the UK National Archives for his work in providing data in a format I could use. His work was above and beyond the call of duty, often outside normal working hours, and is much appreciated. Interpretations of the data, and any errors made in so doing, are my own responsibility. In all my military history research, William Spencer of the National Archives has been a constant source of helpful expert advice. Thanks are also due to Luke Smith (then of the Imperial War Museums) for inviting me in 2012 to take part in OWD. I am very grateful to members of the Zooniverse team, including Chris Lintott (Principal Investigator), Victoria Van Hying (Humanities PI) and Jim O'Donnell (OWD lead developer) for their advice on crowdsourcing projects. Victoria Van Hying was also extremely helpful in the development of this article. Amy Carmichael (Imperial War Museums) is to be thanked for supporting this project, as is Lucy Grayson for reading a draft of the article. Last but by no means least Operation War Diary would not exist at all without its army of dedicated volunteers, the 'Citizen Historians' - thanks are due to them all.

---

<sup>40</sup> Paddy Griffith, *Battle Tactics of the Western Front: The British Army's Art of Attack 1916-18* (New Haven, CT: Yale University Press, 1994), pp. 79-82.