# BADC User Statistics Report 2013

*G A Parton*

## 1. Introduction

The British Atmospheric Data Centre (BADC) came into existence in 1994 to respond to the needs of the Natural Environmental Research Council's (NERC) desire for a dedicated UK data centre for atmospheric research.[1] Originally the Geophysical Data Facility (GDF), operated by the Science and Engineering Research Council (SERC), served less than 200 registered users, from which the BADC's registered user community has now grown to over 22,500 users. During the intervening period the BADC archive has grown to over 1 Pb. of accessible online data and was amalgamated with the NERC Earth Observation Data Centre (NEODC) in 2005 to form the Centre for Environmental Data Archival (CEDA).[2]

This report presents details of the current active user base with a historical review where suitable information was available to the author. The primary sources of information for this review were the user database maintained by CEDA, HTTP and FTP download logs and BADC website access logs.

It is hoped that this historical review will help to provide insights into the BADC user community to enable CEDA to continue to provide improved user services primarily targeted towards its core user community, while also enabling support for an ever diversifying user community.

## 2. Growth of BADC user base

The CEDA user database contains information for registered users for the BADC back to mid-1996 and NEODC[3] from 2005. NEODC figures are included here for information only and 2013 figures are as of March 2013. (Non-registered users are discussed in section 2.2).

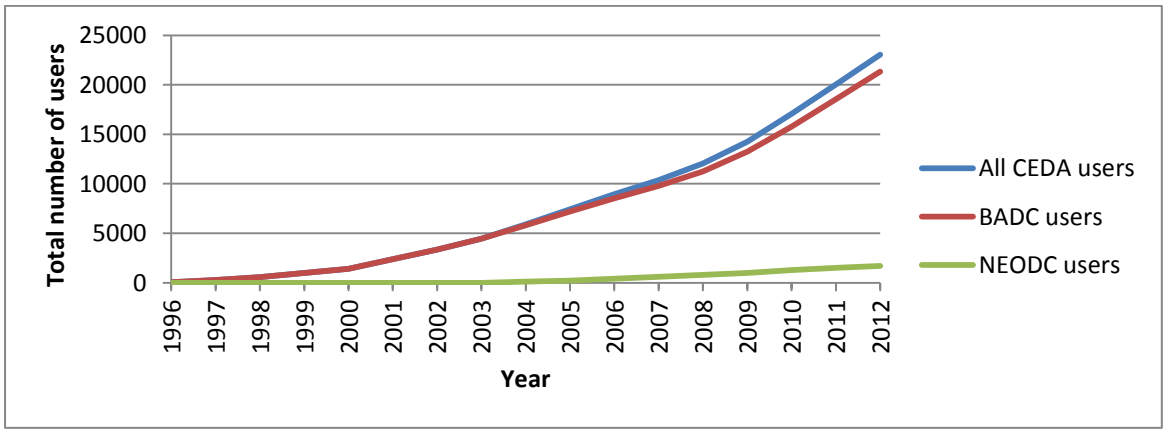


Figure 1: CEDA user base growth

[1] Prior to this the data holdings were managed by the Geophysical Data Facility (GDF) at the Rutherford Laboratory in Oxfordshire, from which the BADC evolved.

[2] For further details of CEDA activities and history, including details of the NEODC, UKSSDC and IPCC-Data Distribution Centre and project activities visit: http://www.ceda.ac.uk

[3] NEODC is one of 4 data centres operated by CEDA, but the only one for which user information is stored alongside BADC in the CEDA user database, hence it's inclusion in this section.

Since 1996 there has been an increase in registered users for the BADC over the past 16 years to around 22,500 at present (Figure 1), with some indication of an increase in the growth rate year on year. A similar pattern of growth in registered NEODC user numbers is also noted.

A closer examination of the numbers of users registering each year (Figure 2) has shown an upward trend from an initial 200+ user in 1996 (source: http://www.ceda.ac.uk/about/history/ ), through to ~500 in 2000 to ~3500 in 2012 for the entire CEDA registered user base, with BADC numbers being slightly below these figures. The increase in numbers registering each year has not, however, been constant, with some years with periods of increase (e.g, 2000-1, 2007-2010) and periods where the growth has plateaued (2001-2,2005-2007,2011-12). The author is not aware of any obvious cause for the variation noted, but it could be speculated that this coincides with specific NERC funded (or similar) research programmes for which participants made greater use of the BADC archive, requiring registration for restricted resources.

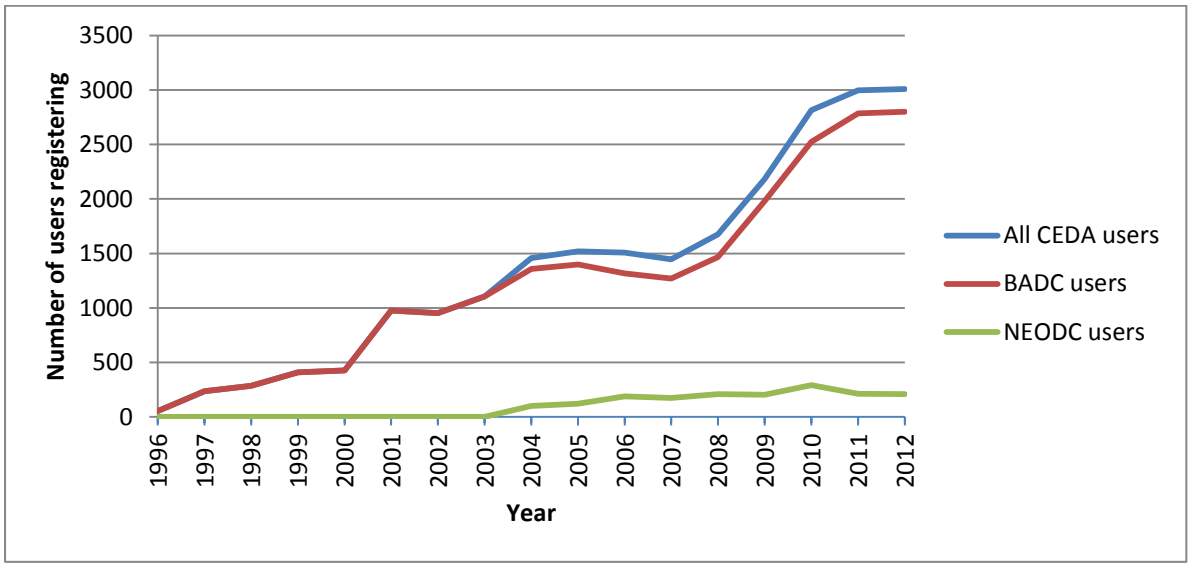


**Figure 2: Number of registrations by year**

There remains, however, a general upward trend in the number of researchers registering with the BADC each year, coinciding with the BADC pulling in users from a larger and increasingly diverse user base (see section 3) and a growing international presence (see section 4). The increase in the wider awareness of the BADC data is as a result of its profile being raised through a variety of means. These include data being referenced and cited in presentations and papers, presentation by CEDA staff at national and international conferences and meetings (e.g. Royal Meteorological Society, EGU, AGU), as well as involvement with projects on the national and international level[4].

There are, however, two caveats that are worth mentioning at this point: the above figures do not show the number of users that continue to use the BADC in subsequent years, nor do they indicate the activity for non-registered users. To some extent these two caveats can be addressed, as discussed below.

[4] For further details of CEDA's projects visit: http://www.ceda.ac.uk/projects

## 2.1 Active users

There is a wide variety in user "usage lifetimes" with the data centre, based on whether the user is a short term user carrying out work for a small project (e.g. a dissertation) or an established professorial grade researcher, who may be an active data centre user for 10+ years. To address this, CEDA uses the number of "active" users to demonstrate the persistent user base from year to year.

An "active user" is defined as one who has access to one or more restricted datasets - i.e. the user must have an account and access to the restricted resource - during the 12 months prior to the reporting period. Due to the structure of the user base the number of "active" users is only available at the time of the value being requested. As of the time of this report this stands at 3909 for all CEDA active users, of which 3608 were BADC users. Comparative figures are available from previous CEDA Annual Reports[5] as shown below in Table 1.

| **Annual Report (April - March)** | **2008/09** | **2009/10** | **2010/11** | **2011/12** |
|---|---|---|---|---|
| Number of active CEDA users in previous 12 months | 2102 (BADC: 1924 NEODC: 178) | Nearly 3000 | 3175 | 3723 |

**Table 1: Active user numbers for period reporting period April - March from CEDA Annual Reports.**

An alternative comparison with previous years' active user numbers can be obtained by examining the HTTP and FTP download statistics for each year, as given in Table 2

| **Year** | **2008** | **2009** | **2010** | **2011** | **2012** |
|---|---|---|---|---|---|
| Number of active users | 2322 | 2799 | 4306 | 4350 | 3497 |

**Table 2: Actively downloading registered users according to http and ftp logs**

CEDA is aware that these download statistics do have some issues, though, as there have been some issues with their consistency, certainly for 2012. In addition, these only record activity via the BADC website and ftp services and do not record other archive service activity offered by CEDA which users may use to access the site.

An independent check is to examine the user logins via the BADC website logs (Table 3). This does not indicate what the user subsequently does, but it is unlikely that a user would log into their BADC account if they never made use of the BADC archive itself. From these logs the number of users for 2012 is 3785, in broad agreement with the download usage logs, thus giving a small degree of confidence in these figures.

| **Year** | **2008** | **2009** | **2010** | **2011** | **2012** |
|---|---|---|---|---|---|
| Active user by web login | 2018 | 2479 | 3116 | 3462 | 3785 |

**Table 3: Number of active users according to web-login**

[5] Available from http:// http://www.ceda.ac.uk/about/ceda-annual-reports/

### 2.2 Non-Registered Users

The information shown so far is for registered BADC users and does not indicate the activity of non-registered users. However, within the BADC download logs there is information about "public" access – i.e. occasions where the user has not logged in with a BADC account id – including the IP address. If the number of IP addresses used by a user is similar for both the registered and non-registered communities (i.e. taking into account users making use of more than point to connect to the internet, e.g. home, office, and similar constraints) then comparing ratios of the number of IP addresses per registered user to the number of IP addresses for non-registered users presents a possible avenue to obtain an estimate for the number of non-registered users. Taking this approach the estimate of the number of users breaks down as given in Table 4:

| | **2008** | **2009** | **2010** | **2011** | **2012** |
|---|---|---|---|---|---|
| Total number of active registered users in period using HTTP or FTP | 2322 | 2799 | 4306 | 4350 | 3497 |
| Estimate of total active un-registered users in period using HTTP or FTP | 1651 | 1917 | 4091 | 3545 | 2462 |
| total number of users in period using HTTP or FTP | 3973 | 4716 | 8397 | 7895 | 5959 |
| Percentage of total active users who are registered | 58 | 59 | 51 | 55 | 59 |
| Percentage of total active users who are not-registered | 42 | 41 | 49 | 45 | 41 |

**Table 4: Breakdown of download statistics used to estimate non-registered users and ratio with registered users**

## 3. Types of Users

During registration for a user account CEDA asks users to provide a number of pieces of information about themselves to help us profile the communities CEDA supports.

As can be seen in Figure 3, the bulk of our users are university based (around 70%), with Government based users coming in at around 15%. While the number of NERC users appears to be relatively low, this may reflect that NERC funded researchers in the National Centres for Atmospheric Sciences and Earth Observation - NCAS and NCEO respectively - whom form our core user community are actually based in universities and not dedicated NERC centres and thus give those as their primary institutions. Overall the break down has shown little variation since around 2002, before which the primary changes were the increase in those outside the prescribed list, with a small decrease in the percentage attributed to NERC centres.

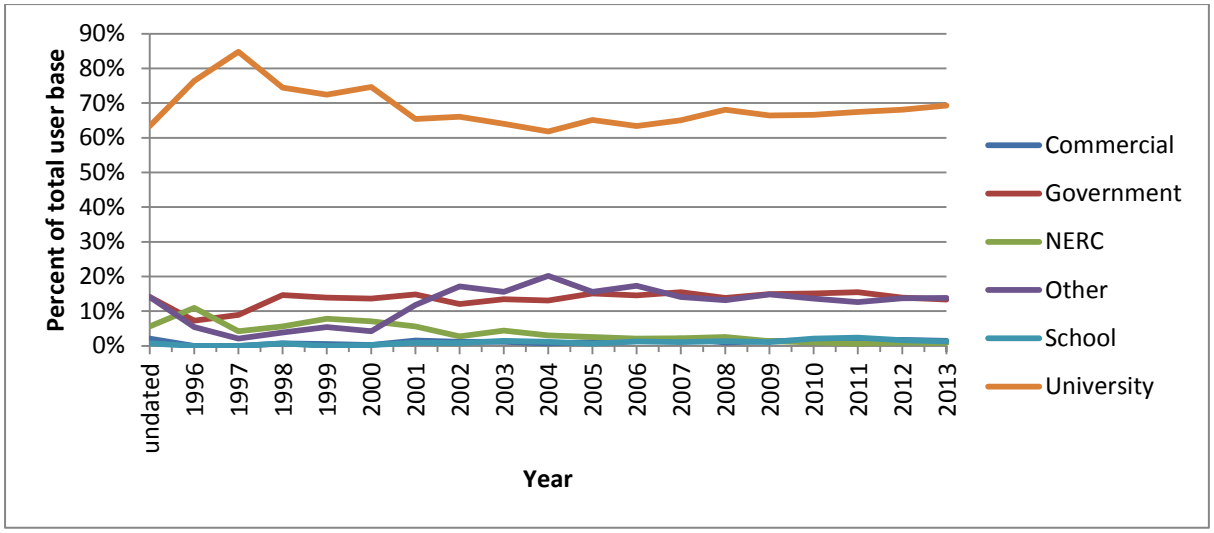


Figure 3: BADC user by affiliation

Figure 4 demonstrates the change in research domains supported by BADC since 1996. Initially nearly all users were classified as atmospheric chemists or physicist, but there has been a steady increase in the proportion of users across all other identifiable subject areas. It should be noted, however, that the "climate change" category was only added during 2012 and while there has been a notable uptake in the use of this category for all new registrations since its introduction, very few existing users have since updated their research domain – indicative that users are unlikely to updated this field in particular of their user details (users do, however, update their titles and institute details far more readily as they review those details during dataset access applications).

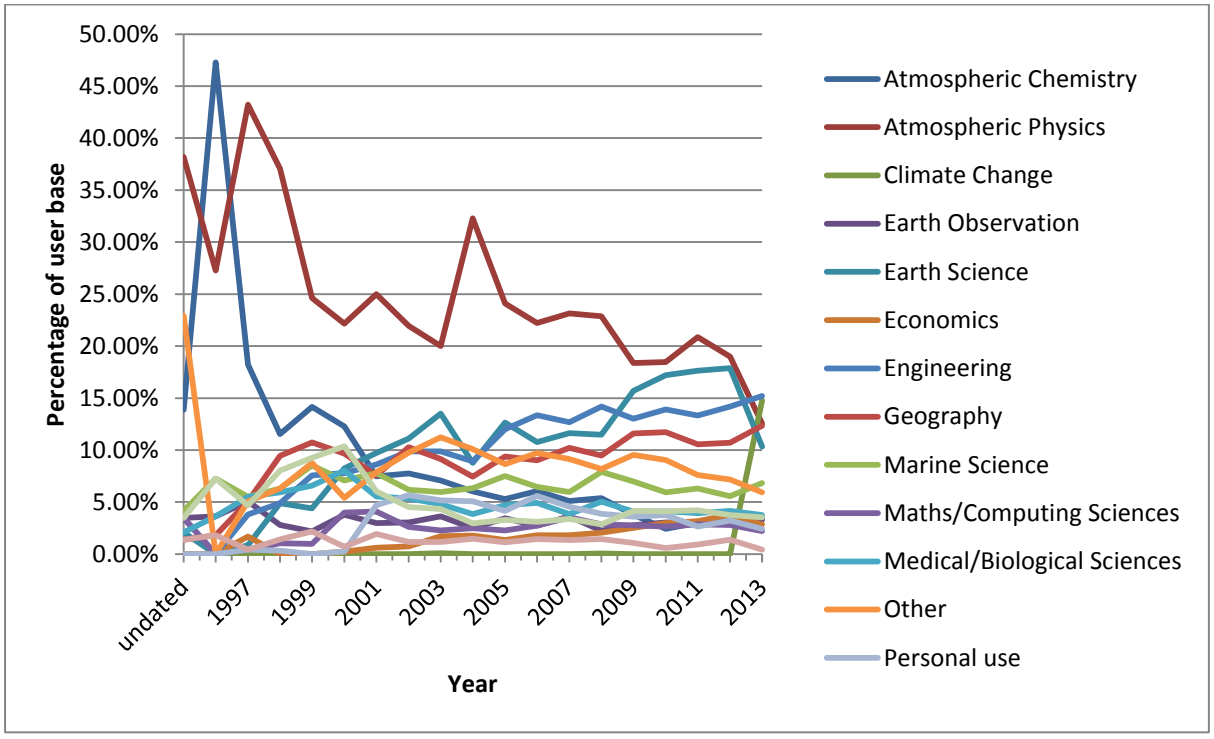


Figure 4: BADC user by subject field

The diversification in the BADC user community demonstrated by the increasing proportions of users from a wider range of fields confirms the experience of CEDA's user support team. This team supply dedicated user support to all BADC users and have noted increasing use of key datasets by researchers using data for a diverse range of research. For example: forensic science (e.g. the decay of evidence based on atmospheric conditions); economics and engineering research into the effectiveness and impact of renewable energy sources; and medical research where weather is an important bias to account for in medical studies.

Taking those users registering for a BADC account in 2012[6] it is possible to also gain some insight into how the user community split by education level, as show in Figure 5. Although the BADC does not store the education level directly, it is possible to infer this information from looking at the user base in a variety of ways. The principal sources of information for these splits were: the degree being studied for, the title of the user and the institute type of the users. Using these fields allows around 75% of the registered users for 2012 to be categorised as follows:

1) "School - no degree": these users were listed as attending school institutes, but not listed as studying for a degree.
2) "School – degree": Initially collating those users that were at institutes listed as schools indicated a higher than expected number of users. On closer examination there were a group that were listed as studying for a degree, thus indicating that these users may be at a "school" but also studying for a degree. This could be explained as either these users confusing the user of "school" as a university level group/department or that these uses may be teaching at a school but still studying for some higher degree.
3) "undergrad": all users at a university institute, listed as studying for an undergraduate degree, but excluding users with Dr or Prof as their title.
4) "postgrad": all university based users studying for Masters of PhD degree, without Dr or Prof as their title
5) "postdoc": all university and NERC centre based users with Dr as their title. This may include true post-doctoral researchers, lecturers, readers etc. without a professorial chair.
6) "prof": all university and NERC centre based users with Prof as their title
7) "other": The above list is not exhaustive and would not cover such users as those outside academic institutes (e.g. a government institution). This number was determined by taking the number of known registrants for each category in 2012 and subtracting the numbers of each of the above user types.

[6] This could also be done for other years, but only 2012 figures have been presented here for breverty.

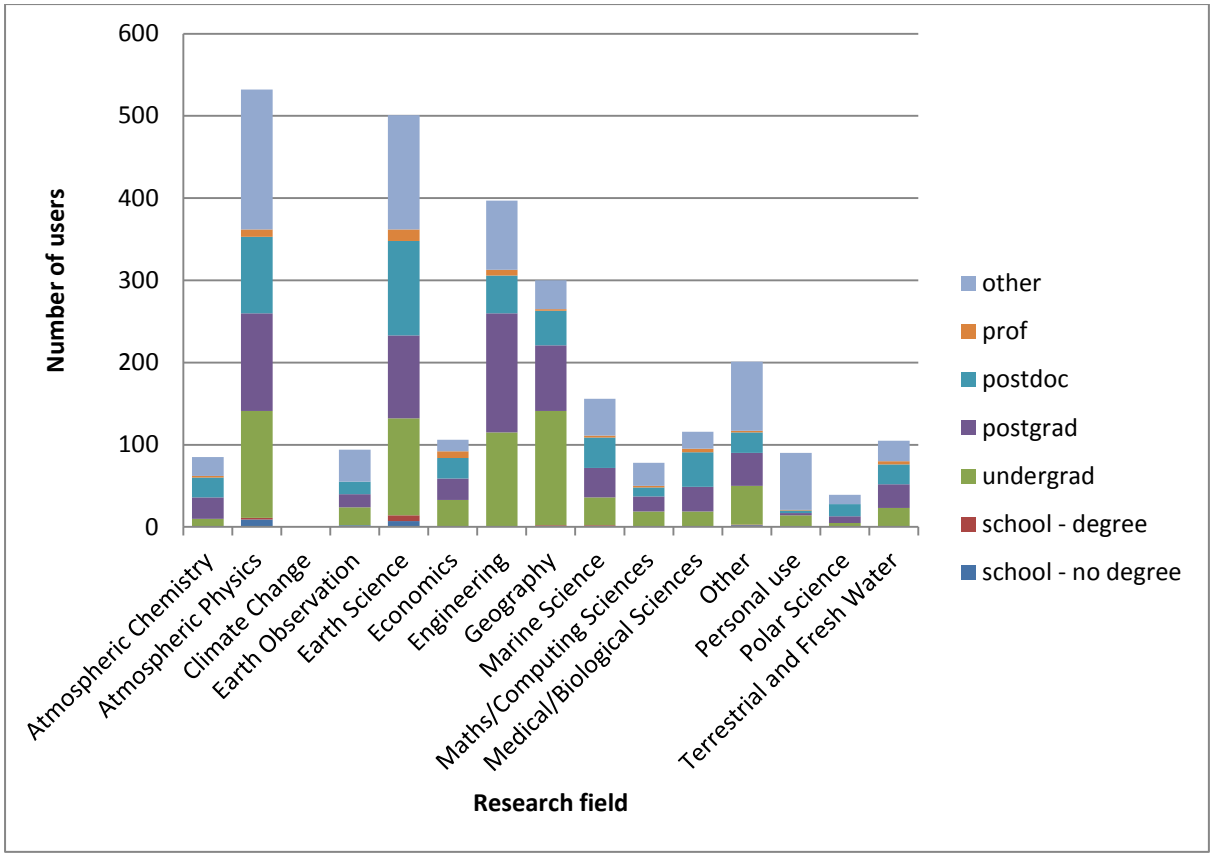


Figure 5: 2012 Subject break down by identifiable user type

As can be seen from Figure 5 there is a reasonably consistent splitting between each of the types of academic users across all subject areas: ~25% are postgraduate, ~25% undergraduate,15-20% are post-doctoral level and 2-3% professorial level.

Looking at the figures for each of these categories of users over time has shown an increase in the number of undergraduate and the small number of school based users. This is evidence of the increasing use of some data for teaching and degree project work – which correlates with an increase in enquiries from such users seen by the CEDA helpdesk. Postgrad, Postdoc and Professorial levels have remained relatively consistent over time, which would concur with career paths of such users. For example, The Royal Society's "The Scientific Century securing our future prosperity"[7] report examined the percentage following various career path choices in and outside of academia with only 0.45% making it through to professorial level, while large percentages of others end up following non-academic careers at some point during their careers (be that in their early, middle or late research career – see figure 1.6 of the Royal Society's report for further information).

## 4. BADC's world wide appeal

Although the BADC was set up to support the UK atmospheric research community it does not preclude users from outside the UK. Since 1996 the proportion of non-UK based users has increased, from initially being nearly 100% UK based in 1996 through to around 65% at present. During this

[7] http://royalsociety.org/policy/publications/2010/scientific-century/

time there has been an increasing spread in the number of countries for which there are one or more users registering to use the BADC, from a handful in 1996 to 88 countries at the time of writing this report. To gain a clearer understanding of any emerging trends in the BADC's global user base these nations have been grouped together either by economic or regional trade blocs - see appendix A for the breakdown of these groupings.

Within the core disciplines of the atmospheric physics, atmospheric chemistry and climate change research communities the bulk of the research is carried out in the UK, USA, Japan, Australia and European counties such as France and Germany. Within the other countries outside this group, the rise of user numbers from the BRIC nations (Brazil, Russia, India and China) is clearly seen in Figure 6. Thus, in recent years the splitting is around 65% are UK based, 10% USA, 10% EU counties (excluding UK) and 10% BRIC nations.

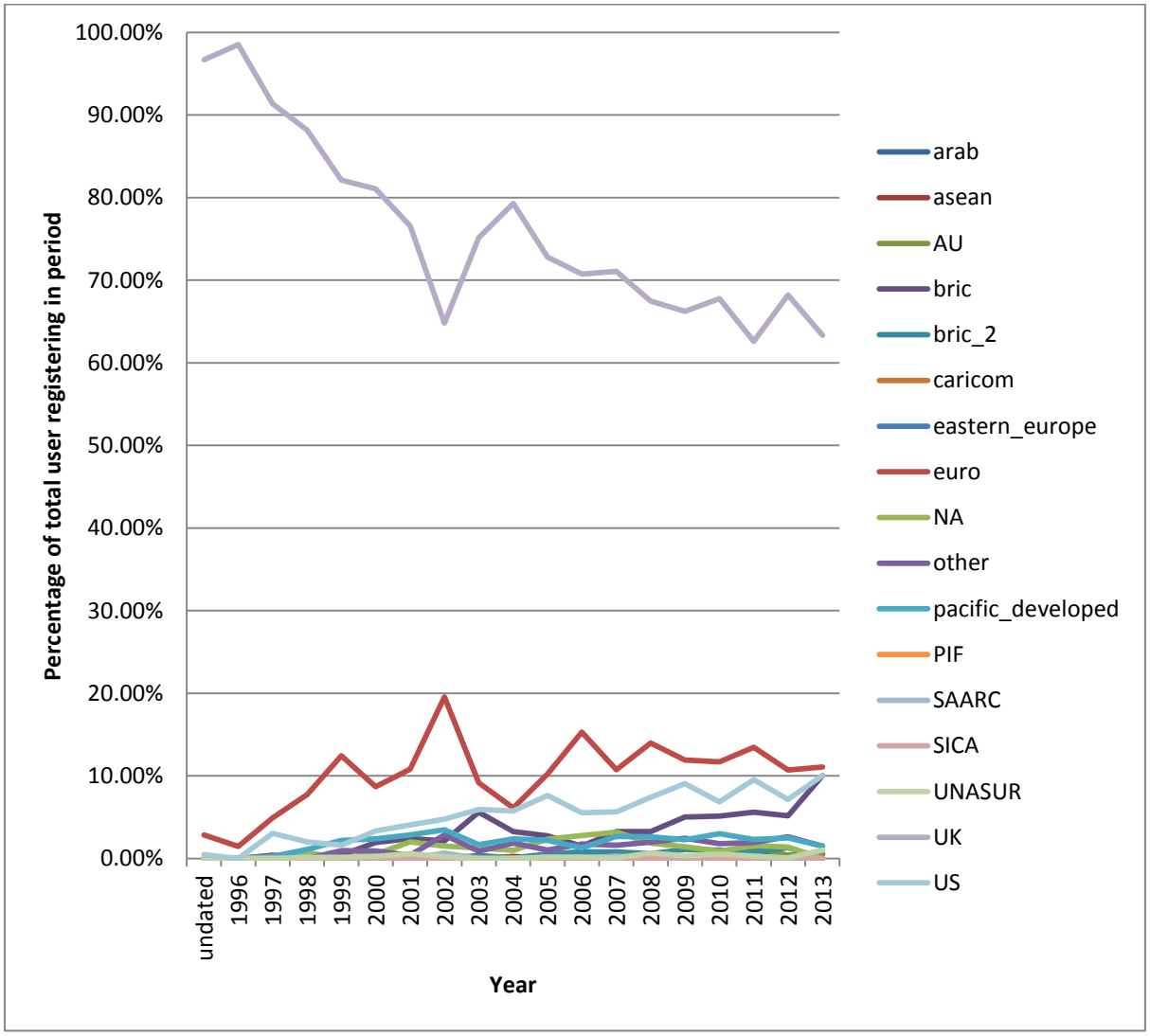


Figure 6: BADC user by economic zone. Please see Appendix A for explanation of the legend.

## 5. Conclusions

This report has examined changes in the user community registering to use the BADC and given some indication to the likely number of non-registered users. While the BADC's primary user community of university based atmospheric and climate change researchers in the UK has remained an important part of the user community, it is clear that the BADC is increasingly being used by a geographically wider and more diverse user community. These findings emphasise that data centres such as the BADC can both facilitate the research of the community it is dedicated to, but also enable research from a wider community too, demonstrating the wider impact of such facilities and the data they hold.

## Appendix A: Regional Trade Bloc Groupings used in this report

To aid analysis of the BADC user community by country the following groupings of nations has been used. These give particular highlight to nations of specific interest to the BADC either as major players in the atmospheric or climate change science research communities (e.g. UK, USA, Euro, Pacific - developed) or of interest from an economics point of view (BRIC, BRIC-2). The remainder are based on the main regional trade blocs as listed under http://en.wikipedia.org/wiki/Trade_bloc ). Countries are only placed into one category to avoid double counting, despite there being obvious overlaps between various categories.

| | |
|---|---|
| **arab** | Countries in the Arab League |
| **asean** | Countries in the Association of Southeast Asian Nations (ASEAN) |
| **AU** | African Union (AU) |
| **bric** | Brazil, Russia, India and China |
| **bric_2** | Countries which are sometimes talked about in relation to the BRIC nations: South Africa, S. Korea, Mexico |
| **caricom** | Caribbean Community (CARICOM) countries |
| **eastern_europe** | Eastern European countries outside of the EU, |
| **euro** | EU countries, excluding UK |
| **NA** | North American Free Trade Agreement (NAFTA), excluding Mexico and US, so basically just Canada (rest of developed North America) |
| **other** | Countries not covered by a trade bloc or of interest to BADC |
| **pacific_developed** | Australia, New Zealand and Japan – these have a prominence in |

| | |
|---|---|
| | BADC user community |
| **PIF** | Pacific Islands Forum (PIF) excluding Australia and New Zealand |
| **SAARC** | South Asian Association for Regional Cooperation (SAARC) – excluding India |
| **SICA** | Central American Integration System (SICA) |
| **UNASUR** | Union of South American Nations (UNASUR) – excluding Brazil |
| **UK** | UK – BADC principal community |
| **US** | US – active in core research fields supported by BADC |