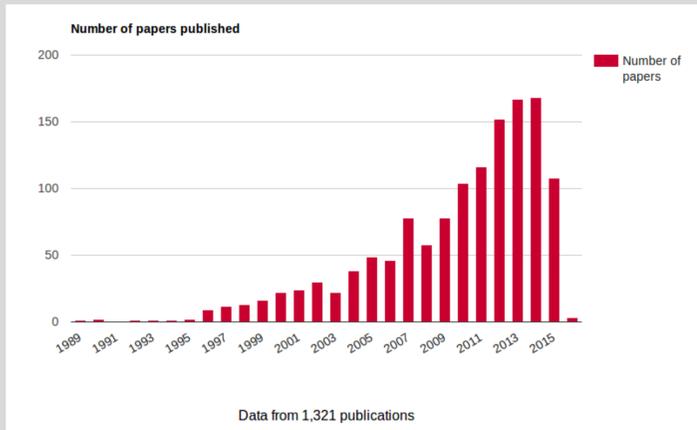
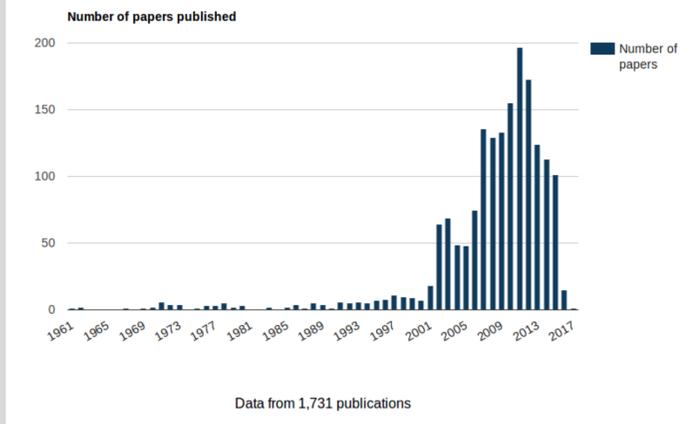


PUMA: First results

Becca Wilson, Olly Butters, Hugh Garner, Tom Burton, Amran Ismail

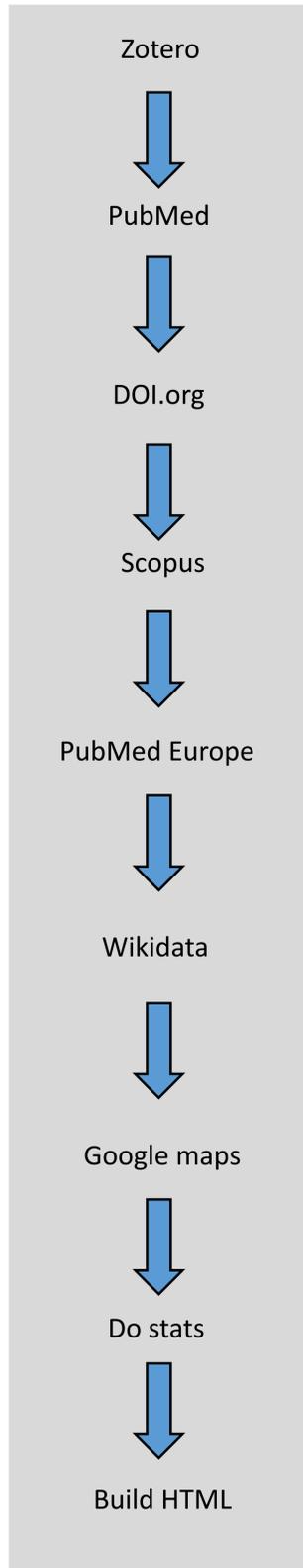


Initial results

The number of publications per year (top left) shows stark differences - the red cohort has a steady exponential growth, whereas the blue one has step changes in it. This highlights the importance of particular data collection sweeps to the blue one's output.

The top level stats for each (right) show some interesting information that would be difficult to calculate manually - total citations for instance.

The locations of the first authors (below) shows a strong trend for English speaking countries. This could feed into decisions such as the worth of building multilingual websites, or having a greater presence at international conferences.



Background

Large data repositories (two large UK birth cohorts in this case) often keep lists of publications generated from their data. We have taken these lists, cleaned them up, and augmented them with as much metadata as we can (see centre column) to turn them into useful data resources in their own right.

We foresee this data as being used for retrospective analysis of project level decisions (e.g. how effective was making change X to policy Y). Additionally we see this as a social science research data set, in the UK birth cohort community at least.

h-index 155 <small>DATA FROM 1,485 PUBLICATIONS</small>	g-index 288 <small>DATA FROM 1,485 PUBLICATIONS</small>	Cohort-Rating 9.759 <small>DATA FROM 1,731 PUBLICATIONS</small>	Mean Citations Per Publication 74.10 <small>DATA FROM 1,485 PUBLICATIONS</small>
c100-index 255 <small>DATA FROM 1,485 PUBLICATIONS</small>	c100-index per Study Year 4.40 <small>DATA FROM 1,485 PUBLICATIONS</small>	Total Publications 1,734 <small>DATA FROM 1,731 PUBLICATIONS</small>	Total Citations 110,040 <small>DATA FROM 1,485 PUBLICATIONS</small>

h-index 113 <small>DATA FROM 1,295 PUBLICATIONS</small>	g-index 200 <small>DATA FROM 1,295 PUBLICATIONS</small>	Cohort-Rating 6.241 <small>DATA FROM 1,321 PUBLICATIONS</small>	Mean Citations Per Publication 49.32 <small>DATA FROM 1,295 PUBLICATIONS</small>
c100-index 130 <small>DATA FROM 1,295 PUBLICATIONS</small>	c100-index per Study Year 2.20 <small>DATA FROM 1,295 PUBLICATIONS</small>	Total Publications 1,322 <small>DATA FROM 1,321 PUBLICATIONS</small>	Total Citations 63,868 <small>DATA FROM 1,295 PUBLICATIONS</small>

Completeness

One of the biggest challenges we have faced is the completeness of the data. Often the initial lists of publications are incomplete or dirty, which requires a huge effort to rectify. Each of the APIs used to then augment the data has to first be linked to a record then the data merged together. This has led to gaps in our exemplar datasets that need to be addressed.

