**Appendix 1. The calculations to determine *X*, the pool of available datasets, at steady state.**

The rate of change in the pool of available datasets *X* is determined by the rate Ps Ys at which datasets Ps are added to the pool by the sharing researchers *Ys*, and the rate at which the pool-size decays (say to a loss in relevance or a maximum in the storage time). So, we write

 (1)

Here the relative decay rate *qx* can be associated with the mean life time of the datasets, i.e. the mean life-time of the publications is given as 1/qx . Also, , is the number of papers produced by a researcher, where *Ts* is the time spent to write it. So, we can write

 (2)

(where Ts is a function of X and Y). If the system is supposed to be at steady state, the change *dtX* is zero, i.e. we can write

 (3A)

i.e.

, (3B)

and as the time to produce a paper is given by (SEE EXPRESSION (3) FROM THE MAIN TEXT)

  (4)

expression (3B) becomes

 (5)

In the right hand side of expression (5) we multiply both numerator and denominator with (1+*f·X*)/(1+*f·X*) leading to

 (6A)

which we can write as

 (6B)

Multiplying out all terms, and rearranging the terms results in the second order polynomial in X,

 (7)

Since for the above top the constant term is negative, i.e. it equals (– Ys ), it is the upper root which specifies the steady state value X for the available pool of data sets, i.e.

 (8)

in which

 (8A)

 (8B)

 (8C)

That is, explicit substitution results in (9)



(which is expression (5) from the MAIN TEXT)

We finally note that according to (6), for fixed parameters, and fixed size Ys, the pool size Xt indeed must converge to the steady state value given by (8) or (9).

 The script to numerically calculate the values of X up until steady state at fixed parameter values is available in the file 'PoolofavData\_v4\_app1.R' via http://hdl.handle.net/10411/20328 V4 [Version].