CrossMark

The Social Life of Data

David N. Kennedy[1]

Did you ever feel like your data are your children? Each of your children is special, valued and loved, no matter what they do. They may sometimes get into trouble (not support your hypothesis), and other times be the head of their class (get you that high profile publication). You were excited when each child was conceived (conceived) and proud when each child was born (collected).

You were very close to your children for their first, formative years (data processing and analysis), but eventually you had to let them spread their wings and go to school (be part of a publication). Once in school, your children met many other children (data contained in other publications); some became friends (could be combined), but others were just temporary acquaintances (could not be combined).

When your children come home from school, you might ask "So, what did you do today?" to which they may reply "Not much" (no one read that paper) or "The class went on a field trip" (the paper was written up in the popular press), or "I got invited to a party" (someone made a data sharing request), or some other reply.

In the 'old' days we just communicated with our children by talking (written request), but today, our communication is really changing. Do you remember the decision to let your child register with Facebook (be listed in a public database[1])? That was a big decision, raising many concerns: Would their identify be safe (private health information be protected)? Would they be bullied (used to show errors in their analysis)? Would they fall in with the 'wrong' crowd (be used to refute their own conclusion)? Would they embarrass you, their parents (turn out to have errors)? But Facebook had its advantages as well. You could see your children's friends (other data they could be combined with); what they 'liked' and were 'liked' by (other results they were pertinent to); and where they visited, electronically (who accessed the data). Your children probably figured out more about cyber security that you knew; how to post public and private messages (communicate availability), how to allow and deny access (set access permissions), how to repost other content (receive attribution for reuse). Today's kids really get around on the social media: Pintrest (dryad[2]), Twitter (dbGaP[3]), Instagram (figshare[4]), (not to mention Tinder) and many, many more sites are constantly getting integrated into your children's web-based social life (are connected to uses and reuses of your data).

Eventually, your children grow up (get archived); and many, but not all, will have children of their own (contribute to a derived data set). You will be a grandparent! You will be proud, and need to share this information with all your friends (get credit for your scientific input to this derived product). If you do not have lots of children and grandchildren to talk about, your friends will not think very highly of you (you will not get tenure).

Children are expensive: all the baby stuff (lab equipment), day care costs (local computing environment), summer camp

✉ David N. Kennedy
David.Kennedy@umassmed.edu

[1] Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, USA

---

[1] https://www.facebook.com/brainhc001/
[2] http://datadryad.org/
[3] http://www.ncbi.nlm.nih.gov/gap
[4] https://figshare.com/

(data management), college tuition (long term archival), etc. Popular children go on to support themselves, but today, many children cannot afford to live independently and remain dependent upon their parents for their entire lives (where is all that old data?). I am sure your parents told you to make sure you were fiscally responsible, and to not bring children into the world that you could not fully support (make sure data sharing support is requested in your grant applications).

One of the first things that children get when they are born is an identification number, perhaps the medical record number from the hospital, and eventually a social security number, at least in the US (does your data have a unique identifier?). This unique identification number is important, and is the key to accessing many services and information sources, and authenticating one's identity. As unique as your child is, there are probably other people around the world with the same name (there can be lots of other data with the same file name). You would be disappointed if your child did not receive their retirement benefits because they could not be uniquely identified as the designated recipient (if you could not get credit for reuse of your data).

So, besides this amusing parallel between data and children, what have we learned? We are reminded to get identifiers for our data, encourage our data to socialize with other data, and plan in advance for the long-term archival of our data. If we can do this in a well-controlled environment, the benefits, individually and as a community, in terms credit, value and reproducibility, are substantial.

In short, if you love your data, set it free, and watch what happens! Its social life may surprise you.