



The Open Dinosaur Project

The ubiquity and increasing respectability of the World Wide Web has dramatically changed the way that science is reported in recent years, especially palaeontology. But so far, it has had much less effect on how it is actually done. While blogs and mailing lists provide increasingly important vehicles for discussing published research, public participation in the scientific process is still mostly after the event: discussing and interpreting (and often disagreeing with) results that have been arrived at by the same rather closed methods that have been in use for many years. Initiatives such as the Audubon Society's Christmas Bird Count, SETI@home (analysing astronomical data for signs of extraterrestrial intelligence) and Folding@home (simulation of protein folding) go some way towards allowing laymen to contribute to scientific endeavours in other fields, but there is nothing analogous in palaeontology; and these projects are in any case very circumscribed in what contribution they allow people to make: provision of raw data and CPU cycles.

Encouraged by the high level of contribution in the comments of our blogs *Sauropod Vertebra Picture of the Week* (<http://svpow.wordpress.com/>) and *The Open Source Paleontologist* (<http://openpaleo.blogspot.com/>), we felt that the time was right to try to change that. In other fields, the technique of "crowdsourcing" – inviting contributions to a project from anyone who feels inclined to pitch in – has been very successful. The most visible example is perhaps the rise of Wikipedia (<http://wikipedia.com/>): that an encyclopedia written and maintained by volunteers exists at all is astonishing enough; but, more than that, its increasing maturity over the last few years has resulted in its becoming a surprisingly reliable resource that is now unquestionably the Internet's primary reference site for most purposes (its coverage of dinosaurs is excellent). But other examples are not hard to find. When the *Guardian* acquired half a million pages of official documents relating to MPs' expenses in June last year, it invited volunteers to review them in search of irregularities, and very quickly uncovered much information that its own journalists would never have had time to sift out (<http://mps-expenses.guardian.co.uk/>). Achievements like these show that the part-time efforts of many untrained people can sometimes be more fruitful than the concentrated focus of a few specialists.

Could such techniques be applied to palaeontological research? Not at such a scale, obviously: far fewer people are seriously interested in dinosaur science than are provoked about being defrauded by their supposed representatives. Still, we had seen enough interest in our blogs, and on mailing lists, to feel that there was a significant body of interested laymen out there who would make a real contribution if the mechanism existed for them to do so. And one evening in Southern California – after a few beers, naturally – Andy and Matt thrashed out a basic structure for such a project, and invited Mike on board the next day. We're interested in how ornithischian dinosaurs evolved from the primitive state of bipedality to quadrupedality – a transition that occurred independently in at least three different lineages (thyreophorans, ceratopsians and ornithopods). To comprehensively study how that happened, and whether it happened in the same way each time, involves analysing measurements of thousands of bones from hundreds of specimens. Much of that information is published but inaccessible, as it is scattered across innumerable papers in various languages; other measurements have never been published in any form. Gathering and collating these measurements seemed like an ideal pilot project for the idea of crowdsourcing science, with the possibility of running further projects in future if the first is successful. And so was born the Open Dinosaur Project, or ODP for short (<http://opendino.wordpress.com/>).



Openness

Our explicit goal is to open up every aspect of the process of producing a formal peer-reviewed publication: to allow anyone who wishes to contribute where appropriate, and at least to be able to see every step of the process. That goal is expressed in the following aspects:

- Most immediately, the project is open to contributions of data – ornithischian limb-bone measurements – whether harvested from published accounts or personally measured. (See below for more detail on how we handle these contributions.)
- Contributors are working on a unified, formatted bibliography and master list of museum abbreviations.
- We are open to suggestions on how to analyse the collated data, and have already benefited greatly from the insights of contributors with more background in statistical methods than we ourselves have. We expect project members to run some of the analyses and prepare figures illustrating trends.
- Although the three of us will prepare the manuscript, we plan to open it to all contributors for pre-submission review, and fully expect the eventual submission to be much the stronger for having gone through this process.
- Crucially, the project will be open to the world, in that the resulting paper will be published in an open-access venue: at present, we expect to submit to *PLoS ONE*, in part because PLoS journals will allow unlimited colour figures and appendices.
- The database being assembled is already freely available, and the snapshot that we perform the analyses on will be included as supplementary information in the published paper. We invite other workers to start using the data as soon as they wish, and request only that they not publish work based on that data before the ODP's own paper comes out, and that they acknowledge the Project for providing the data.
- Finally, the process of assembling the paper is open through blogging (and through the comments that project participants leave on the blog entries). At each stage, we explain what we are doing, why we are doing it that way, and what we expect to do next; and we solicit comments on the future plans. When the paper is published, the ODP blog will be a unique “paper-trail” documenting how it came about.

So far as possible, we want the project to have no secrets from the world.

About the data

Data for the ornithischian gait project comes from two main sources: personal measurements, and published information. While in general members of the public do not have access to the former, it is increasingly possible for anyone who is interested to read the literature. This is partly because of the swing towards open-access publishing, and partly because of digitisation projects that are making older publications freely available – either formal projects, as in the AMNH's digital library (<<http://digitallibrary.amnh.org/>>), or ad-hoc collections such as the O. C. Marsh papers (<<http://sauroposeidon.net/marsh.html>>). One of the most encouraging fringe-benefits of the ODP so far has been seeing how possible it has now become for an enthusiast with no formal affiliation to access the primary literature.



We are alert to the potential for incorrect transcriptions to reduce the quality of the database. Accordingly, we require each measurement to be contributed by two different people before it is accepted into the database. Measurements are publicly released only after having been verified in this way, so as to avoid inadvertently influencing verifiers by allowing them to see the initial contributed value. Measurements that have been submitted only once are noted in the public data as having been submitted, but the values are not given.

Raw data is on the level of individual specimens, and is recorded separately for left and right elements: sternal plate, scapuloacoroid, humerus, ulna, radius, metacarpals, manual phalanges, femur, tibia, fibula, metatarsals and pedal phalanges. For some elements, only the length is taken; for others, more measurements are recorded: for example, femur measurements include length, minimum circumference, midshaft mediolateral width and midshaft anteroposterior width. Of course, most specimens do not include enough preserved material to allow measurements to be made for all relevant elements, so the measurement matrix is very sparse. At the time of writing, one of the most active topics of discussion on the blog is how best to cope with missing data in the various analyses that we intend to run: in this, as in many other matters, the involvement of our contributors puts us in a much better position to apply optimal techniques than we would otherwise be.

At the time of writing, the database contains measurements for 1,659 specimens, of which 507 were personally measured by contributors and the other 1,152 taken from the literature. Of the latter, all but three have been independently verified, and 346 of them double-checked. Aside from personal measurements, the remainder have been harvested from 220 publications. These vary from data-rich papers such as Tumarkin-Deratzian (2009), which contains measurements for 126 specimens, Carrano (2006) (94 measurements) and Maidment *et al.* (2008) (84 measurements), all the way down to the 102 papers that contain measurements for only a single specimen each. It is encouraging to find that the papers that include the most measurements are largely recent ones.

The impressive tally of 1,659 specimens is not quite an accurate representation of the data: some specimens have multiple entries for different bones – for example, because the different bones were described in different publications; in other cases, specimens that have been renumbered or transferred to different collections appear under more than one specimen number. Deduplication is currently under way. Also in progress is the combination of multiple specimens into whole-species records by averaging measurements taken from adult specimens: some analyses will use the per-specimen data and some the aggregated per-species data.

About the contributors

At the time of writing, the data collection phase of the project is essentially complete. Contributions to the database have been made by 46 separate people, all of whom will be given the option to be listed as co-authors on the paper. (Although this is an unusually long authorship list in palaeontology, it is dwarfed by some papers in other sciences: King (2007) showed that the number of papers with 50, 100 and 200 authors is on the increase, with 131 papers published in 2005 having more than 500 authors.) Contributors have widely differing backgrounds, though mostly with some science component: among the most active are professional palaeontologists like Andy Farke and Matt Carrano, but also a database manager for a health-care IT company, a biology student specialising in feather development, a librarian specialising in geology and chemistry, and a camera



salesman who also sculpts dinosaur models for museums. (Interviews with some of these people are available on the ODP website.)

On the negative side, nearly half of the participants have provided measurements for fewer than ten specimens. It's not clear what we could or should have done to encourage more wholehearted participation. We might have established a threshold number of contributions necessary to earn an authorship, but that would seem to go against the principle of openness: all contributions are valuable, after all. If the long tail were longer, then numerous contributors providing just one or two sets of measurements each could be very important on aggregate; as it is, 90% of the data was provided by just the top 13 contributors. A more positive perspective is that these 13 are mostly people who would otherwise not be involved in palaeontology, for whom the ODP presents a unique opportunity.

Into the future

The ODP is not the last word in crowdsourcing palaeontology. In fact, to the best of our knowledge, it is the first that is open to the general public; and our fondest hope is that it will catalyse other efforts that greatly surpass its achievements. Once the paper is published, we will sit down and post-mortem the project: figure out what we did right, what we got wrong, how we might have communicated better or made better use of our volunteers. It's too early to do that at this stage, with the analyses yet to be run and the paper yet to be written, but we do have some sense of what we hope to see happening in the wake of this initial open dinosaur project:

- First, of course, we hope to discover new and interesting insights into the bipedal-quadrupedal transitions within Ornithischia, write them up and illustrate them clearly, and get the paper through peer-review and published.
- We hope that we will find other useful ways to use the limb-bone measurement database, and that other workers will find yet further uses for it beyond those that we have envisioned ourselves.
- If there is appetite for it, we may well follow up the ornithischian gait project with further projects under the ODP banner: perhaps expanding the limb-measurement database to other clades, perhaps tackling completely different problems; hopefully making use again of our existing contributors, and also bringing new people into the fold.
- We want the blog that is being written throughout this project, and the accumulated data-files, to serve as a publicly accessible example of how science happens, what scientists actually do – how data is gathered, collated, analysed and interpreted.
- We hope that non-professional palaeontologists who have dipped their toes into our world by participating in the ODP have learned enough on the journey, and become confident enough, to progress to other projects outside the ODP fold.
- Best of all, we would like to see other palaeontologists learning from our mistakes and successes, and using crowdsourcing in their own work – so that in a few years what we're doing with the ODP will no longer look unusual.



Conclusion

When we launched the Open Dinosaur Project, we did it rather tentatively, not sure how it would be received. We expected pessimism, derision or even hostility from among the ranks of professional palaeontologists, but have been pleasantly surprised at the complete absence of such attitudes. Far from opposing this opening up of science, many professionals have enthusiastically joined in with the data-gathering exercise, and others have offered guidance and good wishes. To be honest, we still do not know for sure that this is going to work; and because everything has been done in the open, we are set up to fail very publicly if we do fail. Still, it seems like a risk worth taking. The ODP is important to us: it's not just about recruiting additional effort into a project, it's about opening up what has traditionally been a rather secretive science, breaking out of the cloisters (Farke *et al.*, 2009). We don't want to work quietly in a dark corner, only to emerge after a few years with findings that we publish in pay-for-access journals that the general public can't read: we want to engage and involve anyone and everyone who cares about science, to bring them into the process and let them share the results.

Michael P. Taylor

Department of Earth Sciences, University College London

Andrew A. Farke

Raymond M. Alf Museum of Paleontology, Claremont, California

Mathew J. Wedel

Western University of Health Sciences, Pomona, California

REFERENCES

- CARRANO, M. T. 2006. Body-size evolution in the Dinosauria. pp. 225–268 in M. T. Carrano, T. J. Gaudin, R. W. Blob and J. R. Wible (eds.), *Amniote Paleobiology: Phylogenetic and Functional Perspectives on the Evolution of Mammals, Birds and Reptiles*. University of Chicago Press.
- FARKE, A. A., TAYLOR, M. P. and WEDEL, M. J. 2009. Public databases offer one solution to mistrust and secrecy. *Nature* **461**, 1053. doi:10.1038/4611053b.
- KING, C. 2007. Multiauthor papers redux: a new peek at new peaks. *ScienceWatch* **18**(6). <http://www.sciencewatch.com/nov-dec2007/sw_nov-dec2007_page1.htm>.
- MAIDMENT, S. C. R., NORMAN, D. B., BARRETT, P. M. and UPCHURCH, P. 2008. Systematics and phylogeny of Stegosauria (Dinosauria: Ornithischia). *Journal of Systematic Palaeontology* **6**, 367–407. doi:10.1017/S1477201908002459.
- TUMARKIN-DERATZIAN, A. R. 2009. Evaluation of long bone surface textures as ontogenetic indicators in centrosaurine ceratopsids. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* **292**(9), 1485–1500. doi:10.1002/ar.20972.