**Deep learning based phenotyping of crop seed banks and herbaria**

**Lead Supervisor:** John Doonan, Aberystwyth University, IBERS
**Email:** John.doonan@aber.ac.uk

**Co-Supervisors:**
Lilian Tang, University of Surrey, Computer Science: h.tang@surrey.ac.uk, Chuan Lu, Dept of Computer Science, Aberystwyth University: cul@aber.ac.uk, Danny Thorogood, AberInnovation Seed Biobank, IBERS, Aberystwyth University: dnt@aber.ac.uk

**Project Description:** Biodiversity underpins long term sustainability of both ecosystems and agriculture. The study of biodiversity, as represented by Herbaria, is one of the oldest branches of biology stretching back at least 400 years.  The physical materials stored within Seed Banks and Herbaria provide a largely unexploited source of information on biodiversity in the recent past.  This project aims to develop automated non-destructive tools, based on state of the art imaging combined with computer vision and deep learning to mine these collections of physical specimens for information that could be used for crop improvement and understanding agri-ecology over the past century

The project is based on the exciting discovery of a huge repository of documented seeds and physical vouchers, assembled over the last 100 years of the breeding programs at Aberystwyth University. These archives are now part of the **IBERS Seed BioBank** located on the new **Aberystwyth Innovation & Enterprise Campus.**  They represent a unique opportunity to explore anthropogenic effects on species and agri-ecosystems in the context of an active crop improvement program.    Current commercial crops have an extremely narrow genetic base selected to perform well in high input farming. This is a general vulnerability in the face of increasing pest

diversity and the need to reduce input to combat climate change. Crop related seed banks include many heritage varieties with potential value for breeding novel resilient strains that can cope better with extreme weather, lower inputs and other pressures arising from future climate scenarios.  The challenge is to extract this information in a systematic manner that useful to researchers, breeders and agriculture.



Fig.1 An example of Machine learning on wheat grains.  Top Left – original image from a bench top micro CT scanner; Bottom Left – the image showing annotation by human expert; Right – accurately annotated 3D uCT model suitable for quantitative measurement of many features.
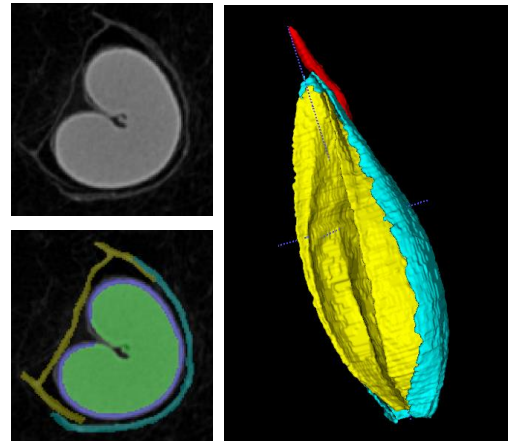
**References:**

Clark, Corney and Tang (2012) Automated Plant Identification Using Artificial Neural Networks", CIBCB2012, IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. doi: 10.1109/CIBCB.2012.6217250

Hughes et al. (2019) µCT trait analysis reveals morphometric differences between domesticated temperate small grain cereals and their wild relatives.  Plant J. 99:98-111. doi: 10.1111/tpj.14312.

Hamidinekoo, et al., (2020) DeepPod: A Convolutional Neural Network Based Quantification of Fruit Number in Arabidopsis. GigaScience in press