| DATA SCIENCE | LINLEY JESSON  WARRICK NELSON |
| --- | --- |
| Outcome More sustainable and profitable systems Impact Indicators Maintained and/or increased crop volumes, value and profitability  New Zealand’s productive environments sustained or enhanced, generating products with verifiable reduced footprints to maintain and/or increase market access Science Targets Whole systems modeling and prediction platform.  Tools developed for eco-verification, footprinting and traceability (water, carbon, greenhouse gases, soil, biodiversity, pesticides, social equity).  Improved production technologies and cultivars based on scaled up knowledge from molecular, physiological, soil, water and environmental science, integrated with system science. Impact Highlight **“Small data, Big impact”**  In PFR, the majority of data sets collected are “small’, of a size that is easily stored within Excel. Thus, these small datasets remain the back-bone of research data collected by PFR, even with the rise of “big data”. Poor management of such small data results in a major a drain on resources: manipulation to enable analysis is time consuming and frequently error-prone, even to the point of making data unusable. At best, this is a waste of effort and at worst is a loss of potential impact or publications. Therefore, it is imperative for the quality of PFR’s science that we have processes in place to ensure the quality of “small data”, particularly that stored in Excel.  To this end, the Data Management in Practice project (DMiP) was initiated in August 2016, and ran for 18 months. The project was supported by Blue Skies money, and the team comprised four biometricians, plus two scientists from outside the Data Science group. We developed several tools, ‘How-to’ documents and guidelines, which are available on the ‘Data Management’ iPlant site (Figure 1).    **Figure 1:** Left: One of the Excel Templates produced by the DMiP team. Right: a page from the Data Management iPlant site.  To introduce and advertise these materials, over the last 10 months we visited every PFR site in New Zealand bar the very smallest. We gave over 50 seminars/workshops, and over 300 people attended (Figure 2) - a marathon effort, but with a gratifying level of attendance, and lots of positive responses.  C:\Users\cflrcb\AppData\Local\Microsoft\Windows\Temporary Internet Files\Content.Word\002.jpg  **Figure 2:** Andrew McLachlan gives a DMiP seminar at Lincoln.  *Selected Comments on the Templates from the Feedback Survey:*  “They are easy to follow and to adapt to different situations. They provide a great base for keeping things organized and standardised, ideal for sharing data but also to simplify analyses”  “We have used the template for all new projects since we completed the workshop. We can see the benefits of it and after the workshop feel confident using it”  “Very well organised, the metadata pages are important and I wouldn't otherwise include them”  The feedback survey plus anecdotal evidence from Data Science group members suggests that the project has led to widespread improvements in the management of small data sets thus also improving the useability of data and efficiency of data management. However, the feedback also indicated that substantial follow up work – modifications to materials, more workshops, continuing reinforcement of ideas- is required. An approach to carry out this work will be devised in the next few months.  The DMiP team (Ruth Butler, Linley Jesson, Andrew McLachlan, Duncan Hedderley, Melanie Davidson and Gareth Hill). OutputsPeer-reviewed Publications Todd J, Poulton J, **Richards K**, Malone L 2018, Effect of orchard management, neighbouring land-use and shelterbelt tree composition on the parasitism of pest leafroller (Lepidoptera: Tortricidae) larvae in kiwifruit orchard shelterbelts. Agriculture, Ecosystem & Environment 260:27-35.  Page-Weir, N. E. M., L. E. Jamieson, A. J. Hawthorne, S. P. Redpath, A. Chhagen, D. E. Hartnett, **L. Guo**, and A. B. Woolf. "Pre-treatment dips to enhance the removal of apple leafcurling midge from apples using high pressure washing."*New Zealand Plant Protection* 70 (2017): 315-315.  R Beatson, **P Jaksons**, K Templeton, 2018. Hops with a difference, Internation Brewing and Distilling Conference proceedings.  John AD Anderson, P. Wright, **P. Jaksons**, A. Puketapu, G. Walker. (2018). Assessment of Tolerance to Zebra Chip in Potato Breeding Lines under Different Insecticide Regimes in New Zealand, 2018. Am. J. of Potato Research (accepted). Client Reports **Butler, R., Jesson, L**., **McLachlan, A., Hedderley, D**., Davidson, M. & Hill, G. 2018. Data Management in Practice: Final Year Report. Plant & Food Research Confidential Report SPTS No. 16123. Internal report 34 p  Currie M, Martin P, Olsson S, Ansorge J, Astill M, Seymour S, Blattmann M, Hunter D, Duffy A, **Guo, L**. March 2018. NT1802 Stage 2 clonal trials Final report on project. A Plant & Food Research report prepared for: Zespri Group Limited. Milestone No. 72258. Contract No. 34362. Job code: P/415802/01. SPTS No. 16131.  NEM Page-Weir, DE Hartnett, AJ Hawthorne, RT Wilkinson, SP Redpath, LE Jamieson, **L Guo**. February 2018. Quantifying the risk of tomato/potato psyllid (TPP) eggs, nymphs or adults on capsicum fruit based on leaf infestation levels. A Plant & Food Research report prepared for: SSIF Vegetables. Milestone No. 69287. Contract No. NA. Job code: P/499000/17. SPTS No. 15996.  **Guo L**, Feng J, Ha B. February 2018. Making the most from Smart Monitoring data for Gold3 kiwifruit (FP1786). A Plant & Food Research report prepared for: Zespri Group Limited. Milestone No. 73095. Contract No. 34574. Job code: P/274126/01. SPTS No. 15920.  Feng J, **Guo L, Alavi M**, Ha B. December 2017. Making the most from Smart Monitoring data for ‘Hayward’ kiwifruit (FP1786). A Plant & Food Research report prepared for: Zespri Group Limited. Milestone No. 73095. Contract No. 34574. Job code: P/274126/01. SPTS No. 15793.  Davidson, M., Watkins, L., Fletcher, J., **Richards, K. & Butler, R**. 2018. Managing YDV in Cereals Sustainably (Sustainable Farming Fund 404939). Summary of Year 2. Plant & Food Research Confidential Report No SPTS No. 16323. A Report Prepared For Foundation for Arable Research.  Gillespie RN, Gosden ML, **Richards KK**, Langer S 2018. Soil nutrient management in dairy farming:a system comparison-year 5 report. Plant & Food Research Confidential Report No SPTS No. 16359. A Report Prepared For Lincoln University.  Morrison SC, Feng L, **Woods CJ**, Andrews FM 2018. Quantification of glucosinolates in swede grown at Gore during the 2017 season. A Plant & Food Research report prepared for PGG Wrightson Seeds Limited. SPTS No. 16008. National Presentations Kui Lin-Wang, Scott Wang, Luigi Falginella, **Lindy Guo**, Guido Cipriani, Vincent Bus, Andrew C. Allan, Richard Espley, Joanna Bowen. WRKY transcription factors (TF) and pathogenesis-related (PR) proteins involved in apple defence following challenge by Venturia inaequalis. MapNet 2017 Capability Developed Catherine McKenzie has been appointed as a Biometrician based at Te Puke. She will start work on 1st July.  Bridget Armstrong began at Lincoln in May as a casual Biometrician, to primarily work with soil and agronomy scientists. She is currently completing her MSc in statistics at Canterbury University.  Carmel Woods attended a two day workshop to become a Software Carpentry Instructor and is now running a weekly Software Carpentry course in Lincoln, with 14 staff on average attending each week | |