

# WheatCAP from a T3 perspective

- The main deliverable of WheatCAP will be to have a significant fraction of USA wheat breeding programs function as a coordinated whole
- It's not so much about work packages and deliverables at milestones but about implementing functions that breeding programs need in coordinated / centralized efficient ways:
  - Genotyping labs
  - UAV image storage and analysis
  - Genomics of wheat for discovery and efficient breeding
  - Workforce training
  - **Data management and analysis**

# Data management & analysis centralization?

- The empirical question that WheatCAP will test:
  - Each of the ~20 programs has its own data solution
  - Will we coordinate / integrate these solutions to the extent that more complex management can be delivered?
    - DNA marker imputation
    - Multi-program genomic predictions
- Unless there is some standardization and joint effort, it will not be possible to deliver these functions

# Data privacy in T3

- There is a public-facing T3:  
[wheat.triticeaetoolbox.org](http://wheat.triticeaetoolbox.org)
- We will set up a password protected WheatCAP private instance of T3  
[wheatcap.triticeaetoolbox.org](http://wheatcap.triticeaetoolbox.org)
- Anyone with the password will be able to access all the data there
- It will be easy to shift data from private to public
- Any data on private older than **24 months** will be shifted to public
  - That is what the grant promised

# Existing breeding program functions in T3

- Experimental design of field trials
- Barcode management for accessions and plots
- Maintaining and displaying pedigrees
- Export and import to the Android Fieldbook mobile data collection
- Seedlot inventory management
- Genotypic data management
- Tissue sample tracking through plates to genotyping outsourcing
- GWAS tool
- Genomic prediction tool

# T3 use as a breeding **process** database

- You all know T3 as a data **warehouse**
  - A place to go look for historical data, say on scab nurseries
- In that capacity T3 does not help you with ongoing breeding tasks
- Three programs have started using T3 for their breeding processes
  - Jessica Rutkoski
  - Mary Guttieri
  - Nick Santantonio (?)
- **3 << 20** There will be growing pains
- Accomplishing breeding tasks on T3 minimizes effort to get data there

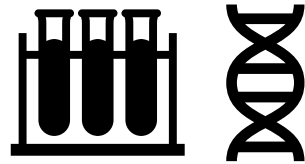
# Genomic prediction

- Field trial phenotypic data on T3
  - Upload accession metadata
  - Either design experiment on T3 or upload the design and field map
  - Either use Android Fieldbook and sync or upload phenotypic data
- Genotypic data on T3
  - Use same accession metadata
  - Genotyping protocols for each low-cost mid-density platform
- Tool to assemble datasets (combination of data types) and predict
- Dataset assembly is an important time sink and will not be done outside of T3 (no XL spreadsheets)

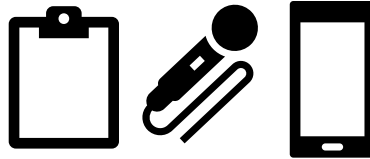
# Getting phenotypes to T3

- Accession identification is a large time sink
- Trait naming is a large time sink
- Formatting spreadsheets to pull data in is a large time sink
- Solution
  - Accession and trait names all in one place
  - Design experiments
  - Collect data digitally
  - Sync with T3
  - A digital ecosystem

# Breeding Cycle “Digital Ecosystem”



Determine genotypes



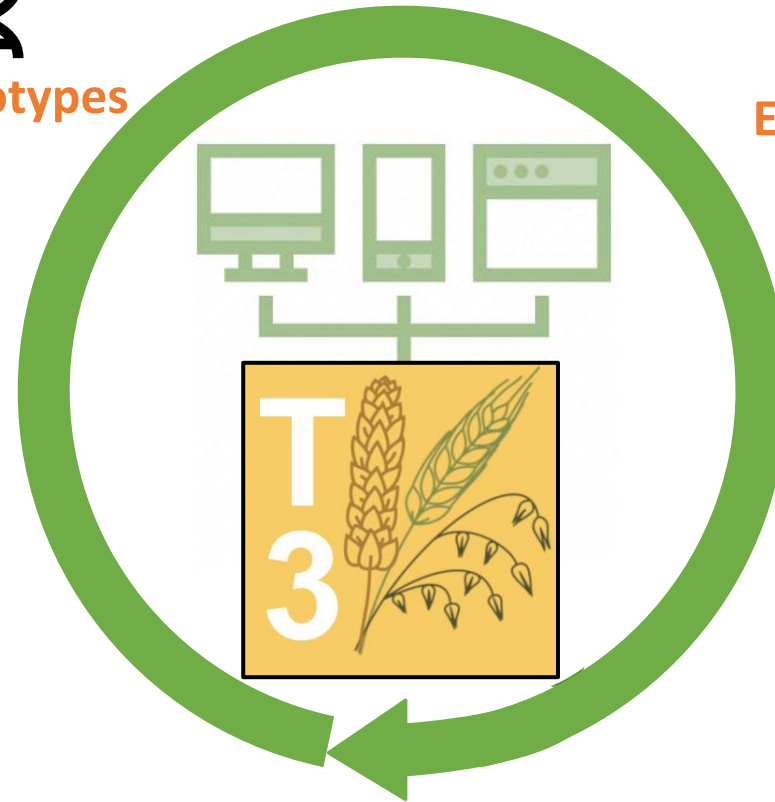
Capture phenotypes



Evaluate genotypes  
Evaluate phenotypes

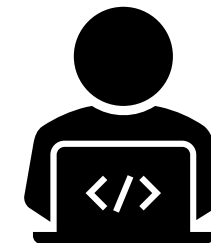


Manage stocks & pedigrees  
Perform crosses  
Design & manage trials  
Manage nurseries



Make:

- ✓ decisions
- ✓ selections (GS, MAS, etc.)
- ✓ reports





# Getting genotypes to T3

- Sample tracking is a large time sink
  - Emails with spreadsheets in different formats linking plates and wells to misspelled accession identifiers: headache
- Use existing accession data on T3 (consistent identifiers!)
- Specify plate designs (multiple plates...) on T3
- T3 sends barcode-identified plate designs to the genotyping labs
- You send barcoded physical plates to the genotyping labs
- Genotyping labs send marker scores to T3
  - All data winds up where we expect it to be with correct identifiers etc.

# Future T3 Tools (speculative)

- PHG marker imputation
- Wheat Breeders User Group: what features do you need?
- Prediction of variance among progeny from a cross: maximize “usefulness” (Jannink)
- Spatial field analysis (Jannink)
- GxE analysis to identify mega-environments for GS training sets (Gutierrez)
- Time-series UAV data analysis for yield relative to maturity (Rutkoski)
- Optimal contributions for polygenic trait maximization while pyramiding (Santantonio)

# Marker imputation

- Low-cost mid-density platforms will have some cross compatibility
- Complete integration (also with historical data) requires imputation
- Imputed Genotyping Protocol
  - All accessions should have this protocol available, regardless of their “raw genotyping protocol”

# Discussion

- To what extent are you willing to change your breeding processes to adopt a “digital ecosystem” so as to facilitate functions that we all want?

