

From Seed Sample to DNA I: Upfront Processing and Sample Tracking

Prepared and presented by:

Dr. Ronald L. Biro

Research Information Management: Laboratory and Automation Systems

Pioneer Hi-Bred International. A DuPont Company

Johnston, IA


- Who is Ron Biro?
- Present high level view of “Sample Tracking” from a data management perspective.
- Laboratory “Sample Tracking” usually involves more than just tracking samples.
 - Materials management
 - Process planning and control
 - Results management
- Will attempt to raise awareness of considerations, issues, and complexity of modern sample tracking systems.

My Terminology

- What is a “**Sample**”?
 - “Experimental Unit”
 - Smallest part requiring a unique ID.
 - Examples: seed, ear, plant, row pool, plot, population.
- What is an “**Analysis**”?
 - Work that needs to be done on a sample.
 - More than one analysis can be run on a sample
 - May require sub-sampling.
 - Usually multi-step process involving extractions, separations, reactions, and measurements.
 - An “**Assay**” is part of an analysis that makes a measurement of a characteristic of the sample.
 - Assays generate “**Results**” and “**Scores**”.
 - Results are usually measurements taken and may or may not require further interpretation (raw data).
 - Scores are interpretations drawn from 1 or more results.
 - Other data can be collected about the analysis process itself.
- What is a “**Job**”?
 - A set of samples requiring the same set of analyses.
 - Usually associated with a **Request** and/or **Project**.
 - Usually scores returned to the requestor as a single report.
- What is a “**Batch**”?
 - A collected set of samples that will undergo identical process steps as a unit.
 - May or may not cross jobs.
 - Allows lab to work more efficiently.
 - Could be physical collection (plate, tray, rack, etc.) or a temporal collection (today’s run).
- What is a “**Test**”?
 - Defined as a unique combination of sample and analysis.
 - Something that will generate a final score.
 - Number of tests is used to estimate effort, cost, and throughput.

Why do sample tracking?



- Reduce human error.
 - Don't rely on human brain to remember everything.
 - Guide user with info for the right material at the right time in the process flow.
 - Unique IDs (serial number or intelligent) for all entities (samples, containers, storage locations, etc.)
- Guarantee/validate/diagnose results.
 - Documents what was done to specific samples. 
 - When, by who?
 - Track reagent lots, quality info, equipment used, etc. for problem diagnosis.
 - Allows association between mother and daughter entities for process backtracking.
- Legal requirements.
 - Regulatory documentation.
 - Support legal dispute resolution.
- Process workflow control/monitoring.
 - Result management vs. Process Control.
 - Assist workers in organizing work to be done.
 - Assist with building day's work (what is available to do, by process step, across all jobs)
 - Guide users through process steps.
 - Support efficiency monitoring.
 - Throughput per unit time for each process step.
 - Failure, re-do rate, etc.
 - Inventory management (ordering).
 - Process modeling.



Sample Processing: Beginning to End

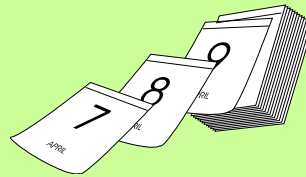
Data/Information management is usually desired or required at:

- Project/experiment planning
- Project setup
- Physical sampling
- Sample preservation
- Sample shipment/storage
- Sample prep/treatment/extraction
- Analysis setup
- Data acquisition
- Data storage
- Data processing
- Data interpretation
- Data reporting

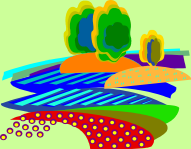
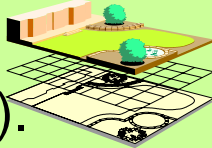


Project/Experiment Planning

– Track the following?

- Who:
 - Requestor
 - Contact list
- What:
 - # of samples needed
 - Types of samples (seed, root, leaf, pollen, etc.)
 - Analyses to be done,
- When:
 - Sampling date
 - Transport timing
 - Scores due date
- Why:
 - Priorities
- Job to job linkage:
 - Samples (or subset of samples) from one job also used in another job.
- “Pre-Assessment”:
 - Sample number estimates – for allocating resources/dataspace.



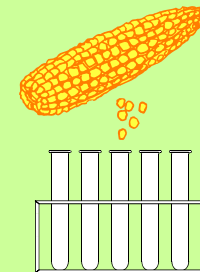
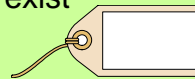
Project/Job/Experiment Setup

- Support Location selection. 
- Support source material layout (field organization). 
- Support Local personnel, equipment, materials allocation.
- Support field operations.
 - Planting.
 - Could be where IDs are assigned/determined (GPS coordinates as an ID?)
 - Tagging. 
 - Field treatments.
 - Chemical treatments.
 - Inoculations. 
 - Phenotype trait collection (associate later with samples).
 - Environment monitoring.



Physical Sample Collection

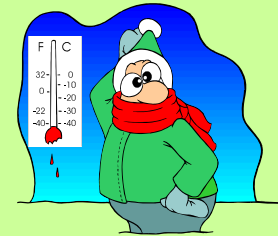
- Provide listing/report of samples to be taken
- Support preparation for sampling
 - Organize equipment, containers, personnel.
- Help find the specific samples to be taken.
 - Link desired “cyber-world” sample to actual physical sample
 - Desired sample may not actually exist
 - “Tags”
 - Tags, stakes, RFID, etc.
 - Order/arrangement from reference.
 - GPS coordinates
 - Row, Range, Plot, Position, etc.
 - Assigned on sampling.
- Monitor physical sampling process
 - Can be labor-intensive.
 - Error-prone/mis-sampling
 - Cross-contamination.
- Support sample containment
 - Associate sample with container or sub-container location.
 - Template arrangement for later added controls and standards.
 - Correct association is frequent point of error.
 - Issue: maintain sample segregation (“sample wander”)



Sample Preservation

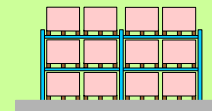
Track the following?

- Reduced temperature preservation:
 - Cooled or frozen?
 - How soon after sampling?
 - Temperature reduced how fast?
 - Was temperature maintained?
- Dry preservation:
 - How soon after sampling?
 - To what moisture content?
 - How dried?
 - Was dry condition maintained?



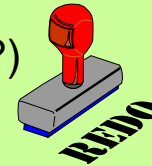
Sample Shipment/Storage

- Support for shipping:
 - Shipping batch.
 - Shipping dates (send and receive).
 - Ship to.
 - Carrier info.
 - Staging area (Is inventory location?).
 - Status (check-out and check-in).
 - Inspection documentation
 - Support reports for problem resolution.
- Support for storage:
 - Inventory location/sub-location: warehouse area, freezer, shelf, etc.
 - Location organization: support arrangement of samples for needed access.
 - Container: box, bin, pallet, etc.
 - Storage period.
 - Storage environment.
 - Status (check-out and check-in).



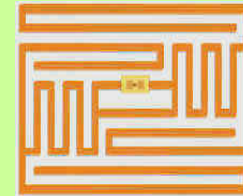
Receive Samples

- Materials check/confirmation.
 - Check-in: Scan to verify?
 - Scan each sample (could be VERY time-consuming)
 - Scan each multi-sample container (assume each sample in container is OK?)
 - Scan each shipment (assume everything is there?)
 - Lost/missing materials.
 - Resample?
 - Same ID or new ID (track original and new?)
 - Cancel samples/tests?
 - Condition/contamination check.
 - Record sample comments (free text vs. predefined entries)?
 - Status change.
 - Date received.
 - Who received.
 - Storage location change.



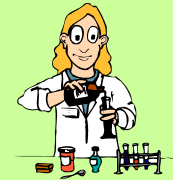
Sample Preparation/Extraction

- Track container change –associate new containers to old
 - Labels
 - Human-readable labels
 - Barcodes
 - RFIDs
 - Label generation
 - When? All at once or when needed?
 - Generate replacement labels.
- Track status change (Ready for next step?).
- Track storage locations of extracts.
- Support analysis/assay method assignment?
- Support quantity/quality determination (Side branch analysis).
 - Track sub-sample – associate sample to sub-sample.
 - Support sub-sample analysis
 - Branching sample tracking process
 - Results used to adjust/normalize sample(s) of main processing flow.
 - Comments.
 - Re-sample/Re-do/Cancel sample tests.



Sample Analysis

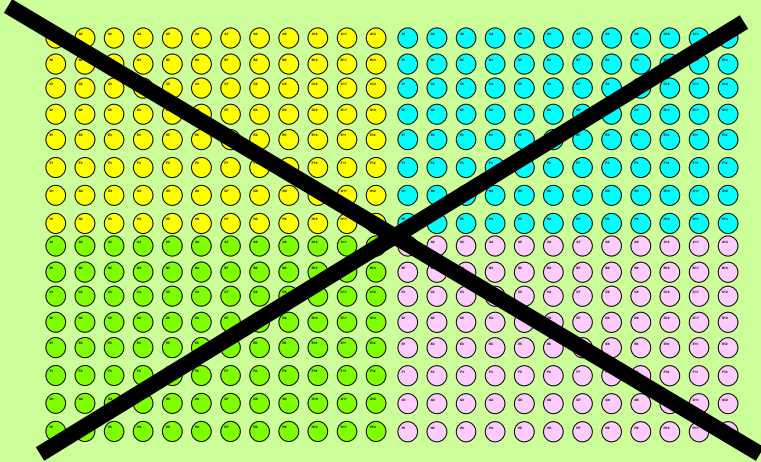
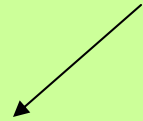
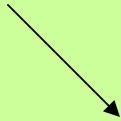
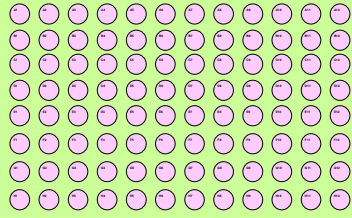
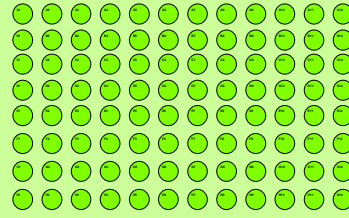
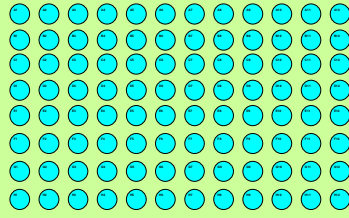
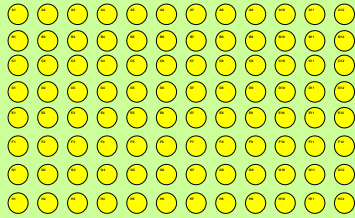
- Support sample/reagent dispensing
 - Assist user in preparing reagents
 - How much reagent?
 - Overage?
 - What reagents?
 - Reagent quality monitoring.
 - Track reagent lots
 - Manual dispense vs. robotic-assisted.
 - Robotics usually don't eliminate all human errors.
 - Manual dispense can be quicker than robotics.
 - Manual is usually less consistent, but robotics can be way off target.
 - Robotic transfer can be easier to confirm/track.
 - Software interface for robotic computer-assisted sample dispensing.
 - How do you talk to the robot?
 - What does the robot need to know?
 - What does the robot assume?
 - What error-handling?
 - Confirm after transfer or before?
 - Dropped/lost/contaminated/mis-dispensed/damaged container issues
 - Re-do, re-work, re-sample, cancel tests, record comments?



Sample Analysis (continued)

- Track container to container transfers.
 - Possibly 1 sample to many analyses. Possibly 1 analysis to many samples.
 - Limited amount of original sample.
 - Above 2 requirements usually have different approaches for efficiency.
 - One method, hard-coded in LIMS, may not meet all requirements.
 - Flexibility leads to complexity!
 - Was it done correctly?
 - Necessary confirmations may place unwanted burden on process.
 - **From low density containers to high density.**
 - Individual samples to racks of samples.
 - 96 well plates to 384 well plates to 1536 well plates.
 - Plates to microarrays.
 - “Mother/daughter” plate and well associations.
 - Interleafed tests difficult to “detangle”.
 - Re-arrayed samples.
 - Arrangement/order of samples is changed.
 - Adds extra overhead to sample tracking (new mother plate arrangement).
 - Data assistance for hit-picking often required.





	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19	A20	A21	A22	A23	A24
B	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24
C	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	C23	C24
D	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	D20	D21	D22	D23	D24
E	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	E17	E18	E19	E20	E21	E22	E23	E24
F	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20	F21	F22	F23	F24
G	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	G20	G21	G22	G23	G24
H	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24
I	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	I21	I22	I23	I24
J	J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	J11	J12	J13	J14	J15	J16	J17	J18	J19	J20	J21	J22	J23	J24
K	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16	K17	K18	K19	K20	K21	K22	K23	K24
L	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16	L17	L18	L19	L20	L21	L22	L23	L24
M	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24
N	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24
O	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12	O13	O14	O15	O16	O17	O18	O19	O20	O21	O22	O23	O24
P	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22	P23	P24

Sample Analysis (continued)

- Track container to container transfers.
 - Possibly 1 sample to many analyses. Possibly 1 analysis to many samples.
 - Limited amount of original sample.
 - Above 2 requirements usually have different approaches for efficiency.
 - One method, hard-coded in LIMS, may not meet all requirements.
 - Flexibility leads to complexity!
 - Was it done correctly?
 - Necessary confirmations may place unwanted burden on process.
 - From low density containers to high density.
 - Individual samples to racks of samples.
 - 96 well plates to 384 well plates to 1536 well plates.
 - Plates to microarrays.
 - “Mother/daughter” plate and well associations.
 - Interleafed tests difficult to “detangle”.
 - **Re-arrayed samples.**
 - Arrangement/order of samples is changed.
 - Adds extra overhead to sample tracking (new mother plate arrangement).
 - Data assistance for hit-picking often required.



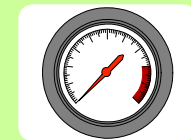
Data Acquisition

- INPUT: Data collection instrument setup
 - Manifest, setup file, sample list, or other files needed to drive instrument. Usually requires file import and/or instrument interface.
 - Data produced and stored automatically or only on command.



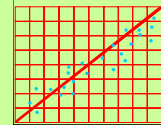
Data Acquisition (continued)

- OUTPUT: Physical world to cyber-world translation
 - Data produced 1 sample at a time or by plate/rack/batch.
 - Relationship of data to samples/analyses/tests may not be stored with data.
 - Current data output formats are highly variable!
 - Delimited text (comma, tab, space, other)
 - Excel ® format
 - Excel can automatically alter values based on Excel data expectations.
 - Proprietary binary formats (instrument-centric world)
 - XML
 - Direct database update
 - Electrical signal: voltage, amperage, or conductivity level
 - Requires hardware intermediate.
 - Where is the data produced/stored locally?
 - Move/import data file info to central database?
 - Associate data with samples/analyses on import (by order of data)?
 - Translate data on import?
 - Ignore invalid data files?
 - Archive raw data files?
 - Manually or automatically archive?
 - What about re-dos (replace or save as unique)?
 - Volume of archived data.
 - How long available for routine access?



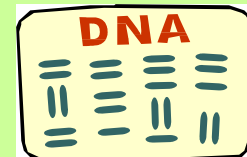
Data Analysis

- Convert raw result data into scores.
- Data translation/normalization.
- Statistical analysis.
 - Comparative statistics.
 - Analysis of variance.
 - Cluster analysis.
 - Probability.
 - Confidence values.
- Automated, manual, or user-assisted scoring.
- Blind vs. informed analysis.
 - Use historical, population, or heritage info to better understand results.
 - Access to: Controls, standards, expected scores.
- Graphical displays of data.
 - Allows user to quickly survey large amounts of data.
 - Can be used to assist in visual clustering, etc.
- Track selected parameters used during scoring?
- Independent or associative scores.
 - Independent: one data point is all that is needed to score.
 - Associative: score is based on comparisons between 2 or more results.
 - Score what subset/superset of collected data?
- Authorize final scores.
 - Draws line in the sand.
 - Closes out a job.
 - Archive scores.



Data Reporting

- How are the scores presented?
 - Format depends on downstream use.
- Include raw results or not?
- Include summary info?
- Include long-term trends?
- Who has access to data/scores?
 - Security issues.
- Use requestor feedback to re-validate scores?
 - “un-authorize” or “re-authorize” scores
 - Overwrite original scores?



Tracking System Components

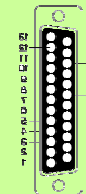
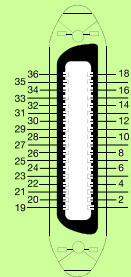
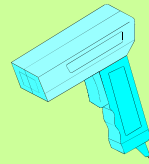
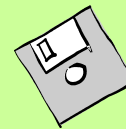
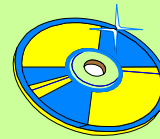
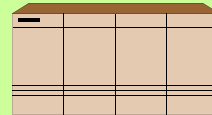
(Software/data complexity)

- Data files (“slow memory”)
 - Instrument setup files
 - Instrument exports (raw data files)
 - Log files
 - Reports
 - Application files
 - Excel ®
 - Graphics files
 - Etc.
- Databases (“fast memory”)
 - Store methods
 - Store reference info
 - Store raw/calculated data
 - Store scores
 - Store entity IDs and associations
 - Store status info
- Operating Systems (hosted by!)
 - Windows, DOS, Linux, Unix, PLC, etc.
 - Compatibility issues
 - Ease of support/upgrade
 - Security issues
- LIMS “Client” Applications (the thinking part!)
 - Process flow control software
 - Project Design software
 - Container tracking software
 - Data acquisition software
 - Detection system software
 - Scoring, data interpretation software
 - Report generators
 - Quality, efficiency monitoring software
 - Quality alert software
 - Control charting software
 - Equipment maintenance software
- Accessory software (call a specialist!)
 - Statistical analysis (SAS, SPlus)
 - Data display systems (Charting packages, SpotFire, etc.)
 - Modeling (ProModel, IThink, etc.)
- Development platforms (the languages)
 - Basic, C, C+, C++ (many versions), MS VS.NET ® (3 versions)
 - Labview, JAVA, Perl, Pascal, Fortran, TCL, Etc., Etc.

Tracking System Components

(Computer Hardware Complexity)

- Workstation computers
- Network servers
- Displays
- Data storage media
 - Hard Drives
 - CD/DVD/Optical
 - Floppy disk
 - Tape
 - Solid-State media
- Printers
 - Human-readable printers
 - Barcode printers
 - Print and apply systems
- Barcode scanners
 - 2D/3D
 - Line vs. Raster
 - Multiple codes
 - Handheld
 - Instrument integrated
- Network/Ethernet components
- Instrument interfaces
 - Serial, USB, Firewire, parallel, ethernet, custom



Tracking System Components (Lab Instruments Complexity)

- Auto-samplers/sample collection equipment
- Sample processing equipment
 - Robotic liquid handlers
 - Extraction equipment/grinders
 - Driers/Reactors/Incubators
- Data acquisition instruments
 - Optical readers
 - Spectrophotometers
 - Fluorometers
 - Single and multi-wavelength
 - Absorption/Turbidity
 - MicroArrays
 - Etc.
 - Chromatography/Electrophoresis
 - Chromatograms
 - Gel images
 - Gravimetric
 - Weighing systems
 - Other technologies
 - DNA sequence
 - Amp plots
 - Conductivity
 - Imaging systems
 - Etc., Etc., Etc.



The Future?

- More samples.
 - Pushed by faster product development.
 - Pushed by more regulatory requirements.
 - Pushed by public desire to know more about products.
- Much more data collected.
 - The ability to collect huge amounts of data will exist.
 - Sifting through the data will be a bigger task than collecting it.
- More process interaction with computers/data.
 - Everything will have a computer in it!
 - Everything will be able to interact with everything.
 - Huge amounts of reference info will be available to support operations.
- Quicker turnaround requirement.
 - Desire for instant results.
- Technologies will change faster.
 - Average life span of a lab method is decreasing.
 - Systems must be able to adapt or must be replaced frequently.
- More pressure to reduce cost.
- More automation.
 - With more samples and more data comes more need to use automation and robotics.
- Move to on-the-fly, on-site analysis.
 - Lab is moved to the field for many tests. How is data transferred?
 - Analyses need to be done very quickly with very little sample.
 - Will require robust and fool-proof systems.
 - The Star Trek ® “tri-quarter” is coming.



Closing Thoughts



- Nature, in her infinite wisdom, has already provided unique ID tags for every living thing. This tag is not just a name or serial number. It provides built-in information on the differences and relationships between the tag's owner and all living things, past and present. Millions of copies of the tag are usually provided to the owner so that copies are readily available to examine. These tags provide the information we need to track and catalog any living thing. All we need to do is to figure out a way to read, interpret, and adapt the tags to meet our desires. The tags are called DNA.