Data Entry: Bane and Boon

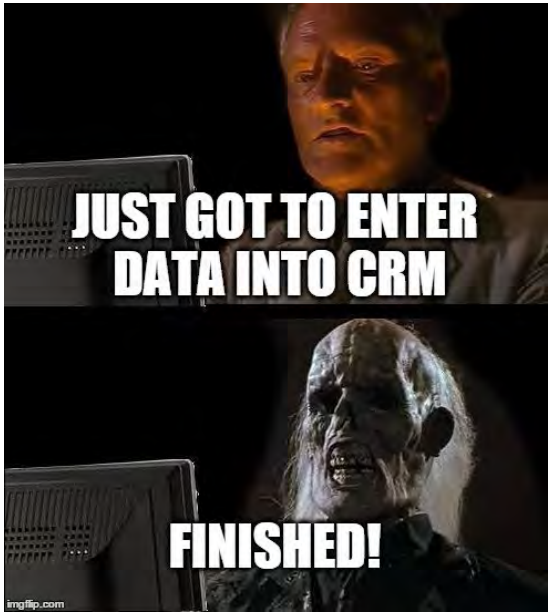
Morgan Taschuk @morgantaschuk

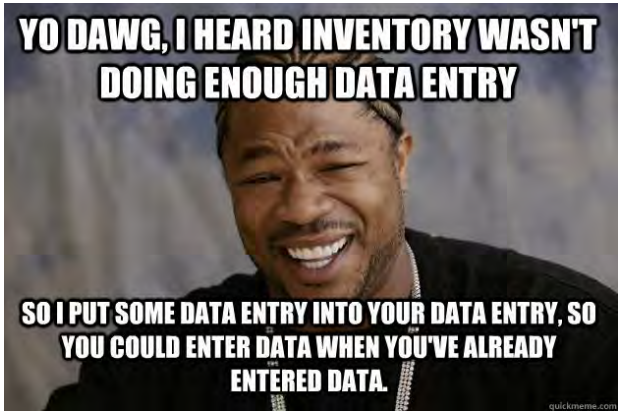
OICR
Ontario Institute for Cancer Research


Prod spa...
110 TiB
Disk Space
100.00%
95.00%
90.00%
85.00%
80.00%
00:00
04:00
08:00
12:00
Percent Production Queue in use
0.15%
0.10%
0.05%
0%
00:00
04:00
08:00
12:00
16:00
20:00
hvmem/memtotal Current: 0%
production
Percent of reads over different index lengths
A) 0 mismatches
B) 1 mismatch
C) Undetermined indices
Percent of reads per lane
Percent1Mismatch
PercentUndetermined
Lane
Cluster
Clusters
350M
300M
250M
200M
150M
DATA
Called Intensity by Cycle
Average Intensity per Cycle
50k
25k
0k
0
50
100
150
200
Cycle
A (Combined)
C (Combined)
G (Combined)
T (Combined)
Base % by Cycle
100
0
Cycle
A (Combined)
C (Combined)
G (Combined)
T (Combined)

https://social.eyeforpharma.com/evidence/real-world-data-collaboration-key-success

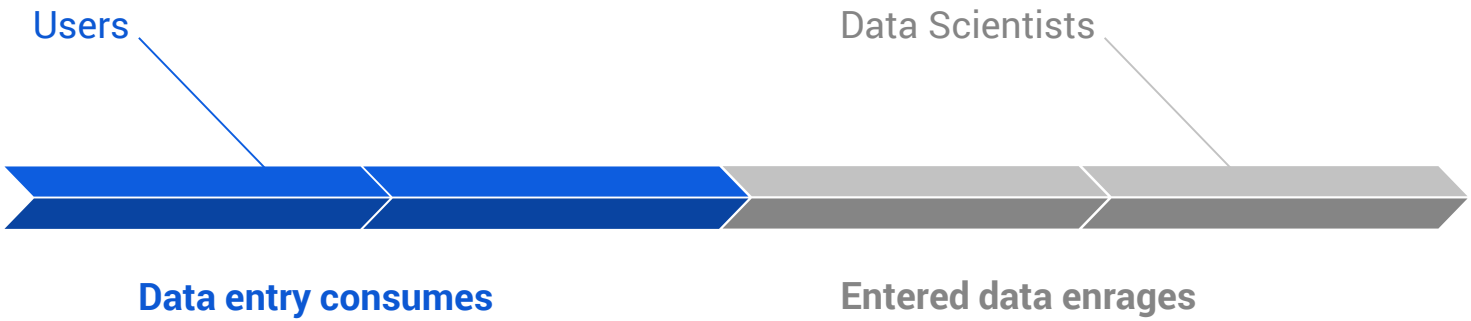
Data Entry


JUST GOT TO ENTER DATA INTO CRM
FINISHED!
imgflip.com


YO DAWG, I HEARD INVENTORY WASN'T DOING ENOUGH DATA ENTRY
SO I PUT SOME DATA ENTRY INTO YOUR DATA ENTRY, SO YOU COULD ENTER DATA WHEN YOU'VE ALREADY ENTERED DATA.
quickmeme.com


DATA ENTRY ERROR?
AIN'T NOBODY GOT TIME FO' DAT!
memegenerator.net

Continuum of data entry misery




ANNALS OF MEDICINE

WHY DOCTORS HATE THEIR COMPUTERS

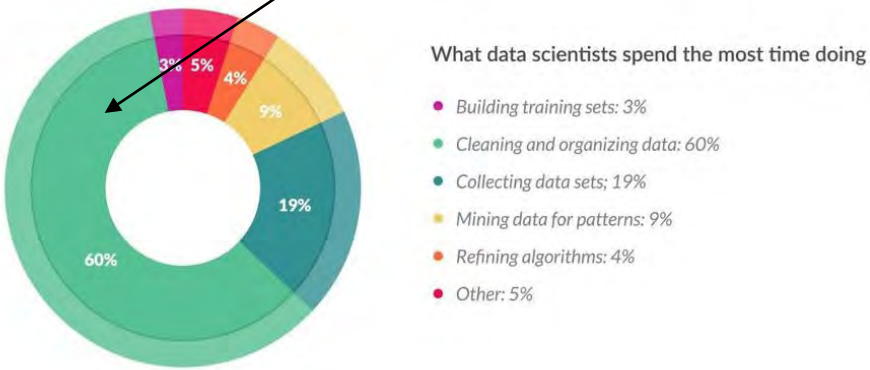
Digitization promises to make medical care easier and more efficient. But are screens coming between doctors and patients?

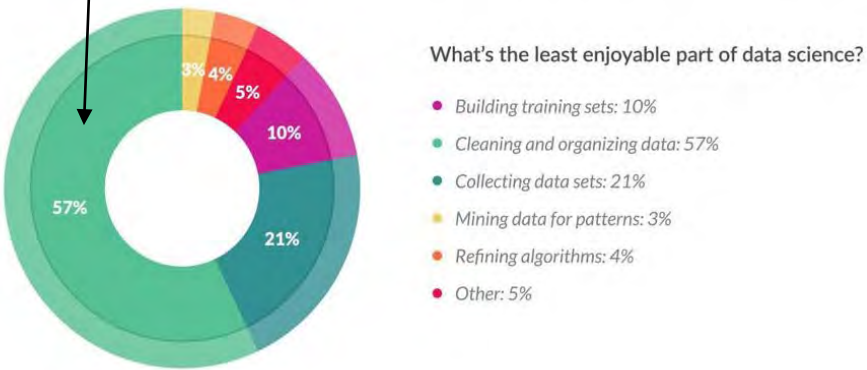
By Atul Gawande November 5, 2018

“people in the medical profession actively, viscerally, volubly hate their computers.”

https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers

Data scientists spend 60% of their time cleaning data and kind of hate it



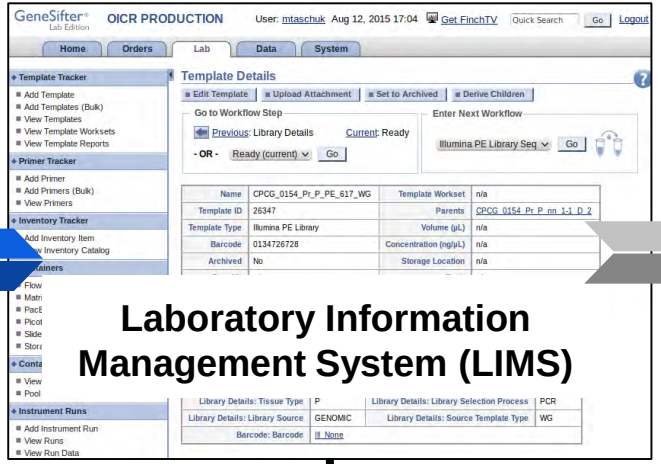


https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

OICR’s data entry continuum

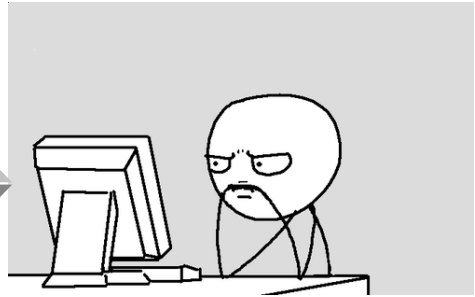
Genomics technicians

Cyanide and Happiness
https://www.youtube.com/watch?v=zTsGcNpr-k0



Laboratory Information Management System (LIMS)

My Team



Users

Data Scientists

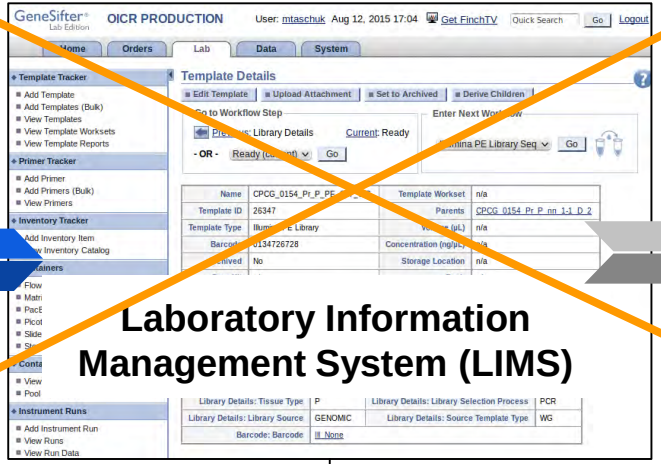
Data entry consumes

Entered data enrages

Seeing the light in 2015

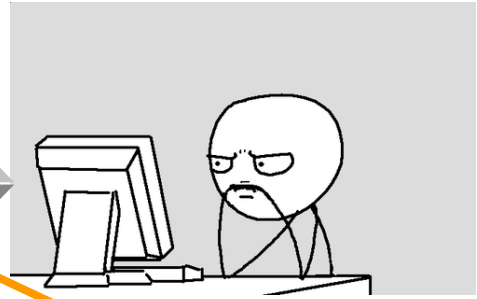
Genomics

https://www.youtube.com/watch?v=zTsGcNpr-k0



Laboratory Information Management System (LIMS)

My Team



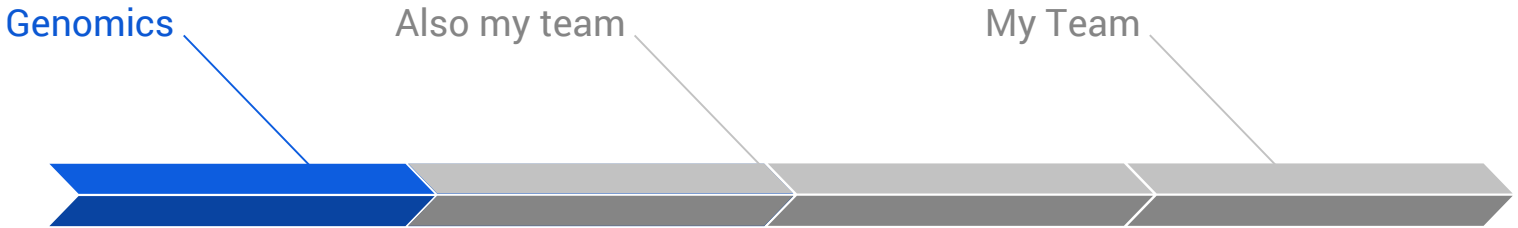
Users

Data Scientists

Data entry consumes

Entered data enrages

Meanwhile, in 2015



Meanwhile, in 2015

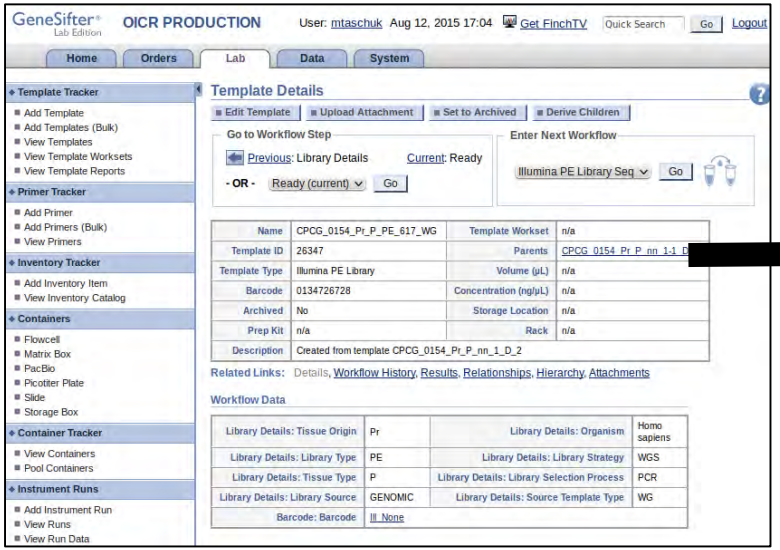
Genomics

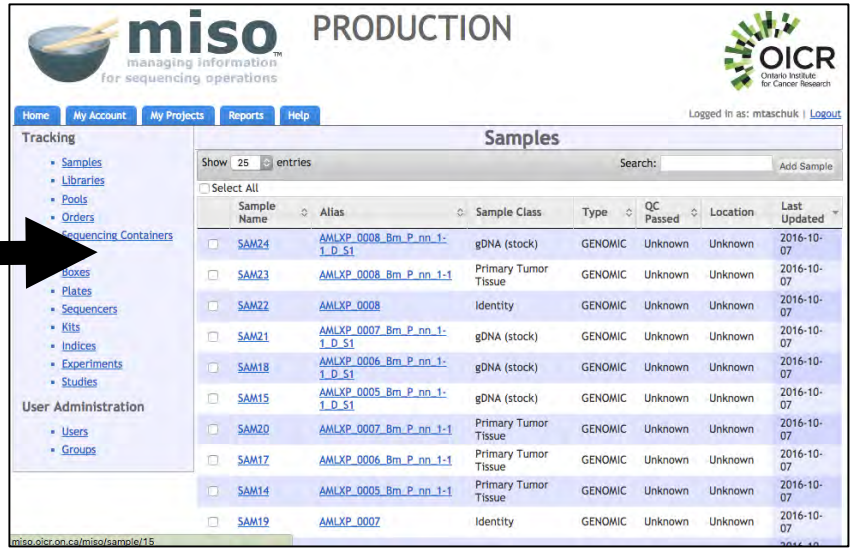
Also my team

My Team

Let’s not repeat our mistakes

Why was the original software so terrible?





Why does software fail?

“Unfortunately, many computer programs that are well designed in terms of technical criteria, such as run-time efficiency, fail to meet the human or organizational needs they were expected to serve.”

Rob Kling, Towards a person-centered computer technology, 1973

doi: 10.1145/800192.805740

MACHINE-CENTERED

1. Efficiency is emphasized.
2. Systems are designed in purely functional terms.
3. Human error is not tolerated. Systems are not error tolerant and provide limited diagnostic information.
4. Users are forced to match the precision required by the machine.
5. Jobs and procedures are designed to simplify machine processing.
6. Human relations are ignored as long as the job gets done.
7. System designs are imposed on users. They initiate but never veto system designs.

Rob Kling, The Organizational Context of User-Centred Software Designs, 1977
doi: 10.2307/249021

Humans are fallible

MACHINE-CENTERED

1. Efficiency is emphasized.
2. Systems are designed in purely functional terms.
3. Human error is not tolerated. Systems are not error tolerant and provide limited diagnostic information.
4. Users are forced to match the precision required by the machine.
5. Jobs and procedures are designed to simplify machine processing.
6. Human relations are ignored as long as the job gets done.
7. System designs are imposed on users. They initiate but never veto system designs.

USER-CENTERED

1. Systems are valued that increase personal competence and pride in work.
2. People are accepted as non-rational and error-prone.
3. Jobs are designed to be personally satisfying. Automated procedures are designed to fit job needs.
4. The burden of precision is placed on the machine. Systems are forgiving.
5. Users easily obtain/create systems that meet their needs.
6. Users can initiate, veto, and collaborate in system designs.
7. Designs and assumptions are intelligible to users through appropriate technique (modular structures) and clear documentation.

Rob Kling, The Organizational Context of User-Centred Software Designs, 1977
doi: 10.2307/249021

Autocorrect

How do we help our users succeed?

miso PRODUCTION

Help | Report a problem | Logged in as: mtaschuk | Logout

Preparation
- Projects
- Samples
- Libraries
- Dilutions
- Worksets
- Pools
- Boxes

Instrument Runs
- Orders
  - All
  - Active
  - Pending
- Sequencing
  - Containers
  - Runs
- Array Scanning
  - Arrays
  - Runs
- Instruments

Tools
- Index Distance

Samples

Create | Edit | Propagate | Print Barcode(s) | Download | Parents | Children | Add QCs | Edit QCs | Add to Workset | Attach Files

Show 25 entries Search:

Name	Alias	Sample Class	Type	QC Passed	Location	Creation Date	Last Modified
SAM201216	PCSI_1073_Ly_R_nn_1-1_D_1	gDNA (aliquot)	GENOMIC	✔	6TH_LIBRARY_INBOX - B02 (647A_H_3_B-3_Lib-Inbox)	2019-05-24	2019-05-24 16:22:21
SAM201215	PCSI_1073_Ly_M_nn_1-1_D_1	gDNA (aliquot)	GENOMIC	✔	6TH_LIBRARY_INBOX - B01 (647A_H_3_B-3_Lib-Inbox)	2019-05-24	2019-05-24 16:22:20
SAM201214	PCSI_1069_Pa_P_nn_1-1_R_1	whole RNA (aliquot)	TRANSCRIPTOMIC	✔	RNA INBOX #2 - E02	2019-05-24	2019-05-24 16:21:19
SAM201213	PCSI_1064_Pa_P_nn_1-2_R_1	whole RNA (aliquot)	TRANSCRIPTOMIC	✔	RNA INBOX #2 - E01	2019-05-24	2019-05-24 16:21:18
SAM201209	SCRM_1040	Identity	TRANSCRIPTOMIC	?	Unknown		2019-05-24 16:02:02
SAM201210	SCRM_1040_Bn_X_nn_1-1	Tissue	TRANSCRIPTOMIC	?	Unknown		2019-05-24 16:02:02
SAM201211	SCRM_1040_Bn_X_nn_1-1_D_S1	cDNA (stock)	TRANSCRIPTOMIC	?	Unknown		2019-05-24 16:02:02
SAM201212	SCRM_1040_Bn_X_nn_1-1_D_1	cDNA (aliquot)	TRANSCRIPTOMIC	?	Unknown		2019-05-24

GOALS

1. Produce clean data

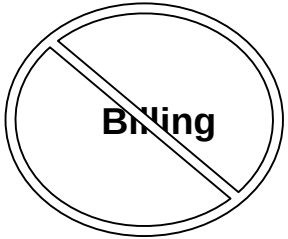
1. Make data entry <u>not suck</u>

MISO Design

1. For who? Technicians
2. For what? Tracking laboratory activities
3. Play nicely with other software
4. Take tips from 40+ years of user-centred design

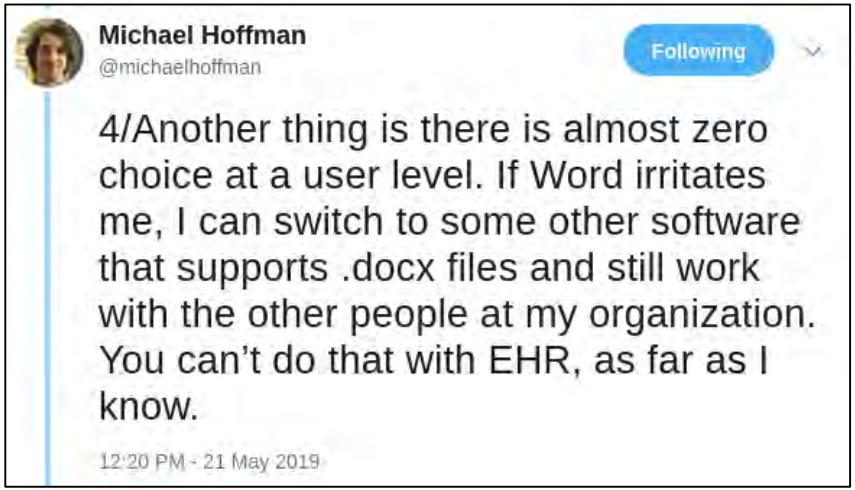
1. Designed for technicians

2. Specifically for laboratory tracking



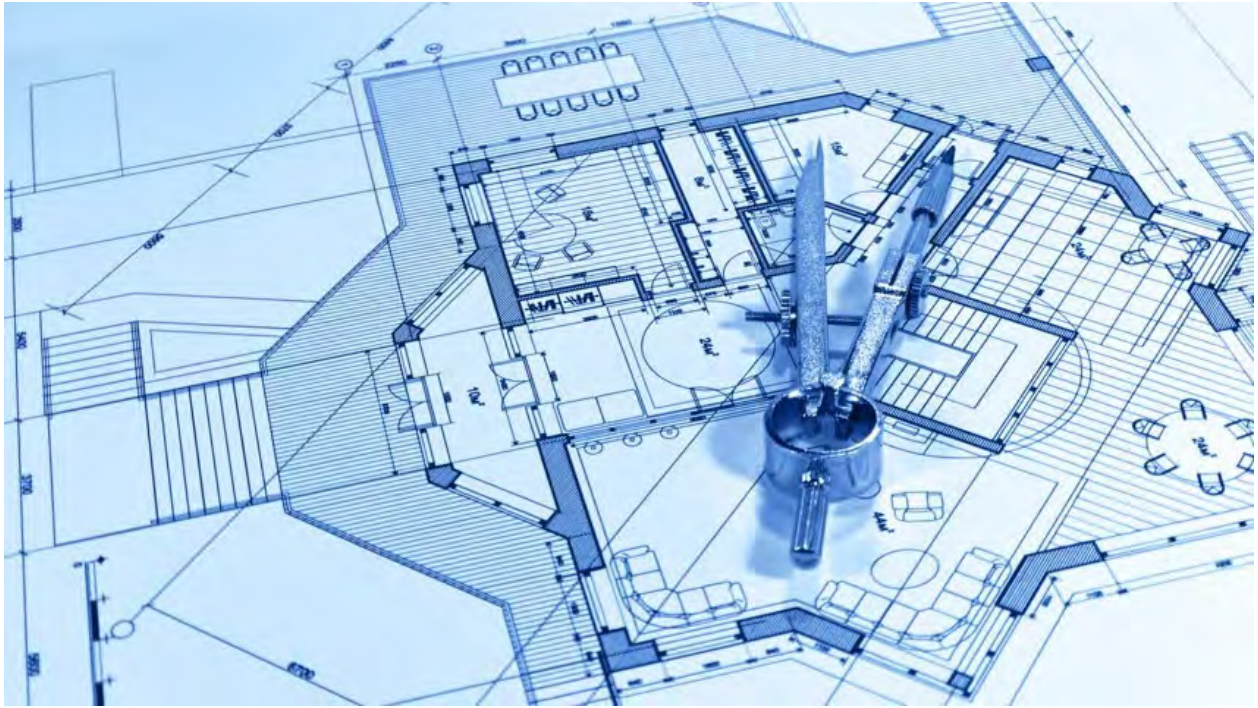
3. Play nicely with other software

- People are going to use Excel -- embrace it
- Let other people extend your software - provide application programming interfaces (APIs)

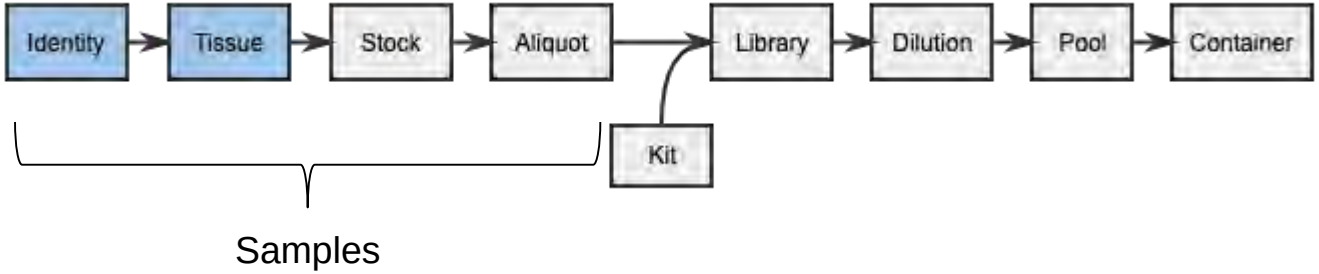


https://twitter.com/michaelhoffman/status/1130870963275522049

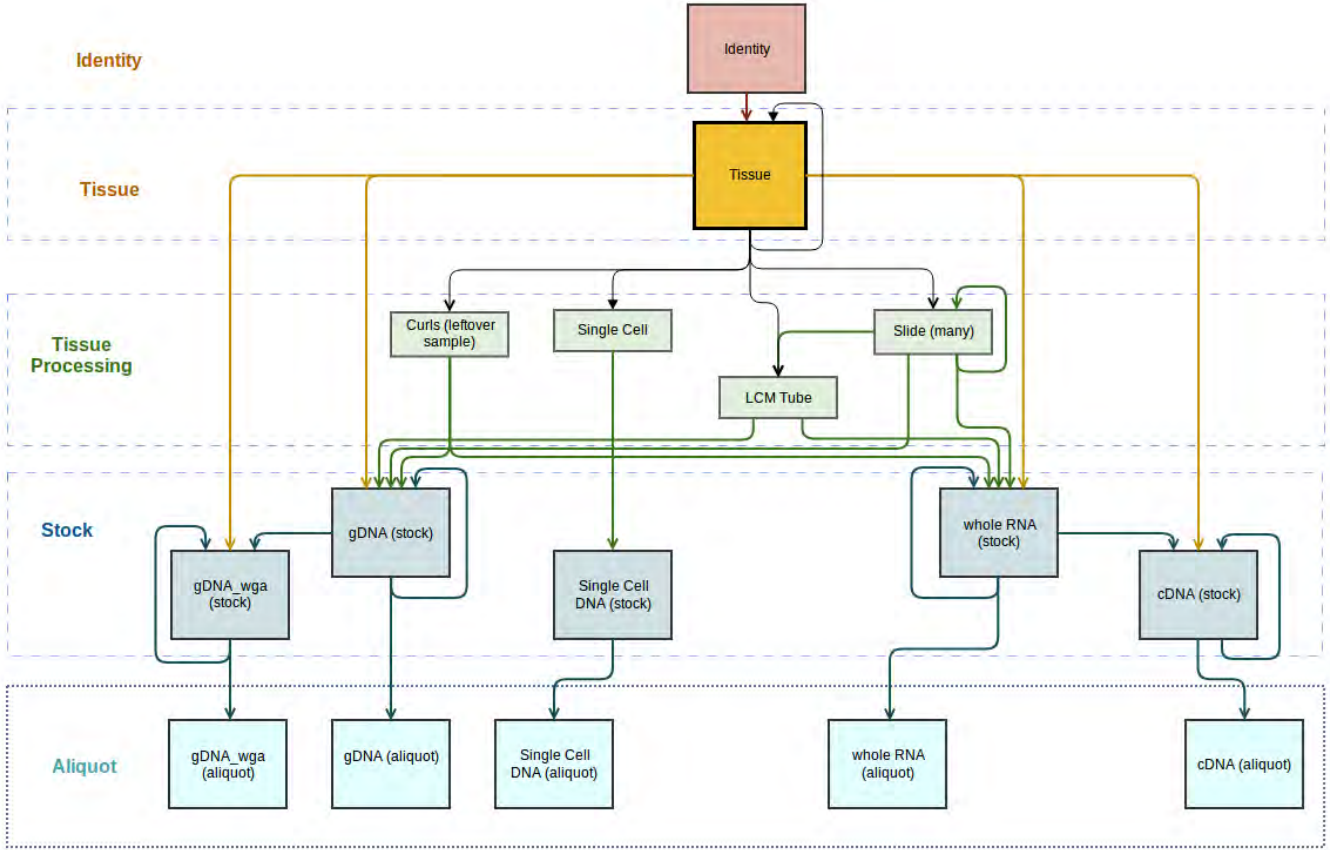
4. Tips from user interface design



Use the language of technicians



Change the software to fit the way they work



Reduce clicks

Doctor, while entering electronic health records

It's 'Oh, who did it?' Why not, by default, think that *I* did it?" She was almost shouting now. "I'm the one putting the order in. Why is it asking me what date, if the patient is in the office today? When do you think this actually happened? It is incredible!" *

1. Shortcuts to common tasks
2. Reasonable defaults (e.g. default to today's date)

* https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers

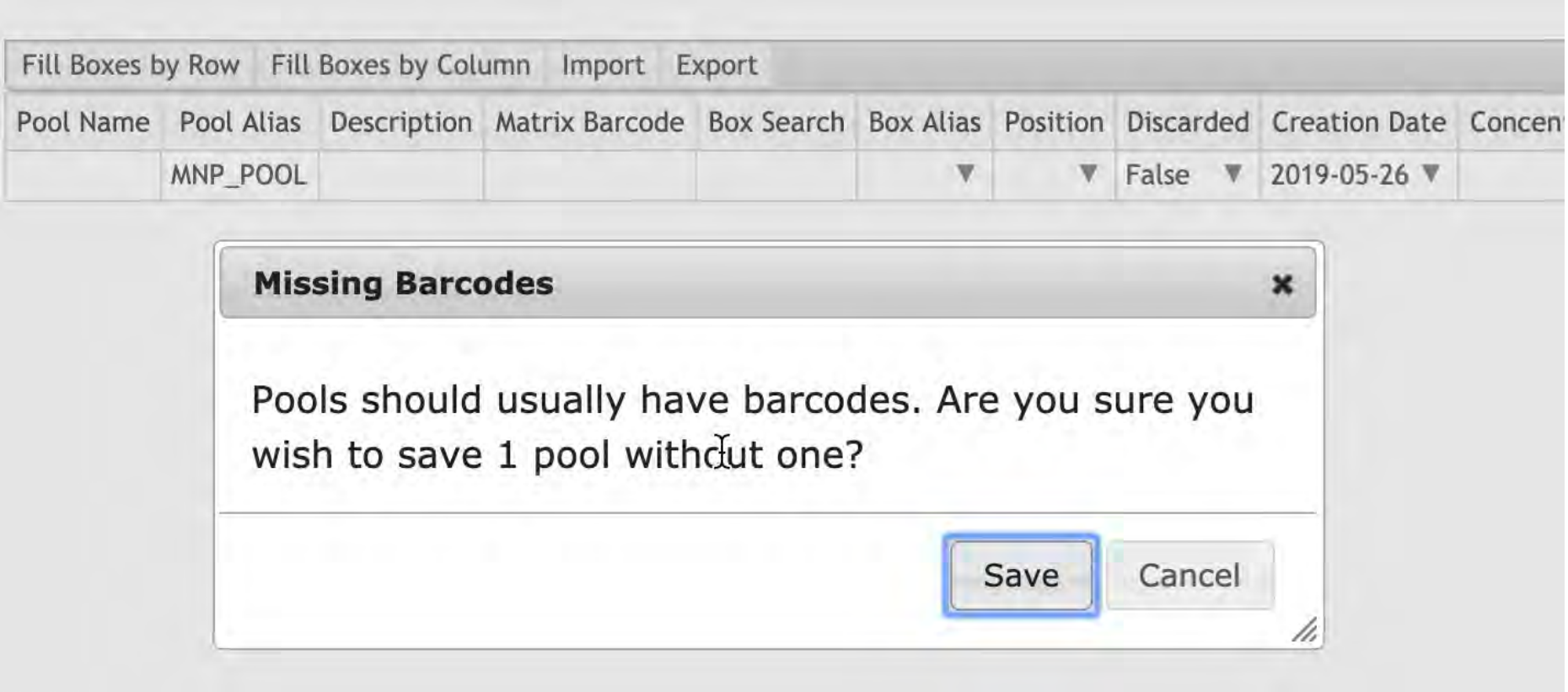
Use alerts *sparingly*

Don’t overwhelm with information

Sort by Sample Location (rows) | Sort by Sample Location (columns) | Fill Boxes by Row | Fill Boxes by Column | Check QCs | Import | Export

Library Name	Library Alias	Sample Alias	Sample Location	Matrix Barcode	Box Search	Box Alias	Position	Discarded	Distributed	Dist
LIB27606	CSUR_0002_Bn_C_PE_370_CH	CSUR_0002_Bn_C_nn_1-1_D_1		0311450980				False	Sent Out	201
LIB27611	CSUR_0007_Hr_C_PE_373_CH	CSUR_0007_Hr_C_nn_1-1_D_1		0311450952				False	Sent Out	201
LIB30444	DCRT_015_Br_T_PE_232_EX	DCRT_015_Br_T_nn_1-1_D_1	6TH_LIBRARY_INBOX F05	0238515764		Seq_Inbox_NovaSeq (BOX799)	H09	False	No	
LIB30445	DCRT_016_Br_P_PE_234_EX	DCRT_016_Br_P_nn_1-1_D_1	6TH_LIBRARY_INBOX F07	0311449220		Seq_Inbox_NovaSeq (BOX799)	B10	False	No	
LIB30447	DCRT_016_Br_T_PE_270_EX	DCRT_016_Br_T_nn_1-1_D_1	6TH_LIBRARY_INBOX F08	0238515753		Seq_Inbox_NovaSeq (BOX799)	G10	False	No	
LIB30448	DCRT_018_Br_P_PE_242_EX	DCRT_018_Br_P_nn_1-1_D_1	6TH_LIBRARY_INBOX F10	0125881713		Seq_Inbox_NovaSeq (BOX799)	H10	False	No	
LIB30449	DCRT_018_Br_R_PE_255_EX	DCRT_018_Br_R_nn_1-1_D_1	6TH_LIBRARY_INBOX F12	0238515726		Seq_Inbox_NovaSeq (BOX799)	A11	False	No	
LIB30450	DCRT_018_Br_T_PE_349_EX	DCRT_018_Br_T_nn_1-1_D_1	6TH_LIBRARY_INBOX F11	0311449203		Seq_Inbox_NovaSeq (BOX799)	F11	False	No	
LIB30451	DCRT_020_Br_P_PE_231_EX	DCRT_020_Br_P_nn_1-1_D_1	6TH_LIBRARY_INBOX G10	0238515725		Seq_Inbox_NovaSeq (BOX799)	G11	False	No	
LIB30452	DCRT_020_Br_R_PE_243_EX	DCRT_020_Br_R_nn_1-1_D_1	6TH_LIBRARY_INBOX G12	0311449013		Seq_Inbox_NovaSeq (BOX799)	H11	False	No	
LIB30453	DCRT_020_Br_T_PE_245_EX	DCRT_020_Br_T_nn_1-1_D_1	6TH_LIBRARY_INBOX G11	0125882362		Seq_Inbox_NovaSeq (BOX799)	B12	False	No	
LIB30454	GLCS_0001_Lv_R_PE_248_EX	GLCS_0001_Lv_R_nn_1-1_D_92		0238515758		Seq_Inbox_NovaSeq (BOX799)	F12	False	No	
LIB30468	MNP_0004_Ad_M_PE_300_WG	MNP_0004_Ad_M_nn_1-1_D_1						False	No	
LIB30469	MNP_0005_Ad_M_PE_300_WG	MNP_0005_Ad_M_nn_1-2_D_1						False	No	
LIB30470	MNP_0006_Ad_M_PE_300_WG	MNP_0006_Ad_M_nn_1-3_D_1						False	No	
LIB30464	PCSI_1068_Lv_M_PE_388_WG	PCSI_1068_Lv_M_nn_1-1_D_1		0311449257		Seq_Inbox_NovaSeq (BOX799)	A06	False	No	

Plan for mistakes and let people know immediately



LISTEN TO YOUR USERS

Bug reports are gifts

Making data entry not suck

1. Plan specifically who and what the software is for (and stick to that)
2. Play nicely with other software
3. Use the language of your users
4. Fit the software to the protocol
5. Reduce clicks
6. Use alerts sparingly
7. Don’t overwhelm with information
8. Plan for mistakes and give feedback immediately
9. Listen to your users

I didn't curse in my head even once!

- One of the greatest compliments we received on MISO

Think of the users

Think of the users

“A person-centred computer technology… can help foster a mature and humane society… If we can not enhance their working environment, what other good can we claim?”

Rob Kling, Towards a Person-Centered Computer Technology, 1973

Thank you

Acknowledgements:

- MISO team for their diligent work and excellent communication skills: Heather Armstrong, Dillan Cooke, Andre Masella, Alexis Varsava
- OICR Genomics, Diagnostic Development, and Translational Genomics Laboratory for their feedback, bug reports, and trust
- Lars Jorgensen and the rest of the Genome Sequence Informatics team

OICR
Ontario Institute
for Cancer Research

Funding provided by the
Government of Ontario.

Ontario