# Designing for Expert Use

Sarah Poon Lawrence Berkeley National Lab sspoon@lbl.gov





## Why UX in the sciences?

The realities of scientific work

- Not a sequential, well defined process
- Supporting artifacts and contexts of use are not always captured by software
- Collaborations might have existing complex software stacks that need to be considered

UX research can help uncover these considerations and UX design ensures that software that take these factors into account.

Adapted from slide courtesy of Lavanya Ramakrishnan



Figures Courtesy: Paramvir Dehal, KBase team





Source: Ameriflux project

## We work as interdisciplinary teams to develop scientific software



- We work closely with application scientist to develop methods and tools to manage the data and workflows
- We use UX methods to understand user needs and convert that into concrete and actionable outputs

Courtesy of Lavanya Ramakrishnan

Developing the "right thing" is important for user adoption of new software, and the UX research and design process is essential to achieve this.

## **UX Design Process**



# Takeaways for Expert Use Design

Understand Work Processes and Goals

SNFactory originally used a search interface as their primary data system. What they actually needed was a tracking system.

SNTrak 6.1	SNTrak 6.1
Nearby Supernov	va Factory SN Tracking System ("SNTrak") v6.1
No Frames.)	
Add a Candidate	
List These SNe (once you have selec	ted the parameters of the display, below, and the supernovae you want to show, further below.)
Which database: snfactory	Additional information (where available)
Sort by: Name Name	<ul> <li>Finding charts for this supernova (link)</li> <li>Usobel's spectrum page for this candidate (link)</li> </ul>
□ Include inline subtractions	All miscellaneous extra information.
Select Candidates by P	riority
List all candidates with screening pr	riority greater than or equal to: 2
List all candidates with follow up pr	iority greater than or equal to: 2
Discoverer Name:	
From Date: (Enter de	stas using format: MM/DD/VVVV)
To Date:	ites using format. MiteDD (1111)
Include supernovae from sets:	
□ \$2001 (Spring 2001)	
S2001 (Spring 2001)	
□ F2001 (Fall 2001)	
□ \$2002 (Spring 2002) □ \$2002 (Fall 2002)	
S2003 (Spring 2003)	
□ F2003 (Fall 2003) □ S2004 (Spring 2004)	
□ F2004 (Fall 2004)	
SNF2005 (2005)	



The original search style interface didn't capture the need to track the currently active candidate supernovae.

- Interviews and observations can reveal underlying workflows.
- Pay attention to "shadow systems" such as notebooks and files.

"If I'm going to go to all the trouble of coding my own plugin, why would I bother using your framework?"

Understand User Motivations and Values

- There can be a tension between the time and effort needed to write code and the value you get being able to run that code at scale.
- We decided to create a visual plugin builder. This helps support exploratory work needed to understand data change and also eases some of the coding burden of developing plugins. The output script can run at scale.

Deduce	About Documentation
Files to Compare	Comparison Queue
Base New	base_file.csv new_file.csv analysis
Load Files	var 4 — var 4a new - base
Comparison Parameters (optional)	
Inspect File: v5_7_0/6838/spPlate-6838-56429.fits v6_0_0/6838/spPlate-6838-56429.fits	
Which HDUs do you want to compare?	
No         Name         HDU Type         Dimension         Content Type           I         0         PRIMARY         Image         (4645, 1000)         spectrum           I         NAR         Image         (4645, 1000)	
Analysis Type	
Name Comparison Type Visualization Type	
PRIMARY array1 histogram	
to Comparison Queue	
	RUN COMPARISONS

Deduce helps users understand data change across datasets. This visual plugin builder helps users define and explore data change.



- Expert users often want to be able to override automations and tweak outputs from algorithms
- If possible, consider human in the loop scenarios before writing the algorithm



The scheduling algorithm didn't take into account the desire to be able to tweak the schedule afterward.



"You had no way of telling that if I try to open this file, whether it will open in a fraction of a second or whether it will take five minutes, because it is running off to tape and doing this thing for me."

Algorithms and abstractions are useful but can't be black boxes. Expert users want to know what to expect.



Accessing data from archival storage takes longer than data on disk. The design shouldn't abstract away this type of information.



- Scientific users want to be able to do their jobs efficiently and without errors.
- Efficiency looks different depending on the purpose and goals at hand
- UI level takeaways for the data overview pattern:
  - **Design for data density.** Minimize what needs to be held in working memory to make a decision.
  - Don't display more data than is needed. Improve signal to noise by hiding ancillary information in secondary layers.

8	Show/Hide Fie	lds 🗦 \Xi Filte	r 6 total strategies					🖶 Print	🛓 Export
	Service Territory	Threat Scenario	Infrastructure Name	Resilience Strategy	Assumed Lifespan	Discount Rate	Annual Avoided Utility Costs	Annual Avoided Customer Costs	Annual Other Avoided Costs
~	AP&T	Tsunami	Substation	Tree trimming/ vegetation management	20 years	10%	\$50,000	\$30,000	\$10,00
~	AP&T	Tsunami	Substation	Underground the transmission line	15 years	10%	\$50,000	\$30,000	\$10,00
~	AP&T	Tsunami	Substation	Upgrading poles & structures with stronger, more robust materials	5 years	10%	\$50,000	\$30,000	\$10,00
~	AP&T	Tsunami	Substation2	Tree trimming/ vegetation management	20 years	10%	\$50,000	\$30,000	\$10,00
~	AP&T	Tsunami	Substation2	Underground the transmission line	15 years	10%	\$50,000	\$30,000	\$10,00
~	AP&T	Tsunami	Substation2	Upgrading poles & structures with stronger, more robust materials	5 years	10%	\$50,000	\$30,000	\$10,00
	Portfolio Benefit 1.1 Portfolio Bene \$1.8	t / Cost Ratio 2 (fit (\$) (PV) <b>3M</b>	Poi	ttolio Benefits (PV, Millio 80.2 (11%) Avoided Util Avoided Cut 35%) Other Avoided	ity Costs stomer Costs ed Costs	\$0.3 \$0.2 \$0.1 \$0.0 \$0.1 \$0.0 \$0.2 \$0.3 1 2 3	A 5 6 7 8 9 Total Benefits	10 11 12 13 14 15 Total Costs — Ne	16 17 18 19 20 t Benefit
	Portfolio Cost (\$) (PV)			Portfolio Costs (PV, Millions)			Average Portfolio Wide Indices		
Portfolio Cost (\$) (PV)			\$0.2 (13%) \$0.3 Capital Inves	stment	8 6 4				

Example of a data overview pattern

## STRUDEL: Scientific software Research for User experience, Design, Engagement, and Learning

Towards developing a comprehensive UX framework for the sciences.

## Motivation: user interfaces across the sciences have many common parts



How can we empower scientific users to develop their own UIs while leveraging our learnings of how to create good scientific UIs?

## STRUDEL: Develop a design system for scientific software



#### **FRONTIER - Scenario Comparison**



#### **Science Software Characteristics**



#### **Design System**

A collection of reusable components that can be assembled to build a UI

#### Components



Design System Guidelines & Implementation

## Towards Generalized Page Layout Templates



ESS-DIN	DATA PORTALS PROJECTS GET STARTED AE	SUBMIT DATA
Search ©	DATASETS 1 TO 25 OF 659 Hote Map >	
Search phrase Q	2 3 27 Ned Botty Most reart ~	
Filter by:		1997 - STA
✓ 쓸 Project	Bahner G.; Ju Y.; Arend K.; Morin T.; Rey-Banchez G.; Wrighton K.; Wila J (2019); Nethane and CO2 chamber fluxees and porewater concentrations US-OWC Ameritius wetland site, 2015-2018. Ameritius Managament	
Project name Q	Q Project, ESB-DIVE repository, Dataset, doi:10.15485/1568865 0  Project, ESB-DIVE repository, Dataset, doi:10.15485/1568865	
> @ Identifier	Goldman A E ; Chu R K ; Denczak R E ; Daly R A ; Delgado D ; Fansler S ; Forbes B ; Banayburu Caruso V A ;	2 / S / C / S / C / S
> Q Region description	Graham E B ; Laan M ; McCall M L ; McKever S ; Patel K F ; Res H ; Remeria L ; Resch C T ; Rod K A ; Ttaly M ; 3 2 1 3 18 2 Talic N ; Torgeson J M ; Toyoda J G ; Wells J ; Wighton K C ; Bregen J C ; WHONDRS Consortium T (2005):	
>	WHONDRS Summer 2019 Sampling Campaign: Global River Corridor Sediment FTICR-MS, Dissolved Organic 1 4 5 10 2 Carbon, Aerobic Resolution, Elevental Composition, and Grain Size. River Corridor and Watershed	
> 🔟 Year	Biogeochemistry SPA Worldwide Hydrobiogeochemistry Observation Network for Dynamic River Systems	
> Access		2 26 1 26 4 1 9 1
	Phas M; Lich HY; Janva D; Bit K; Nong W VC; Stopki S (\$122); LICM-based altergrand bianess charged dist of inpiction matter for the lowers. Ref-weekdord: Ethinated of Carbo Bok et Carage in Trapical Freeds (Stor In Carbon Chara) (Stor Inc. ESB OFIC: sponkry; Dataset, act to Statistical Data	$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$
	Pares 3, Bayle 10,2022; D-bable extendenabilities of each or spatially datas. States Theorem 2014 Each Development or spatial and the spatial states and the spa	
	Prese Bi, 1666 Tree C, 1667 Tree M (2002) C Assess them includes and California system and with set attends a systematic in the hybridesta assessions. Return Quark Biogendermitry Source From Area, BSS-000 monology, Dataset, doi:10.1546/1186071	
	Soong J : Hicks Pries C : Castanha C : Pontas R : Olki N : Tern M : Schmidt M : Riley W (2022): Soil C stock and	Replaced shortputs Map data 80000 1000 km samaar 2 Terms of Us
	inersci 🎎 @Energy 🕮	

• Other stands and and the stand	0 tites	selected			Download ~		
Institution     Q. unama     Institution     Institu	6 'mest r	relevant" results shown + 0 additional results found		Show Most Relevant Results O Show All Results			
Notestime         Nature (Notestime)         Nature (Notestim	Eve	erything v Q corona		Store Filters 🛱 👌			ł <sub>o</sub>
Protocol	^	Metagenome-Draft Assembly ZJ_Prado_40		Number of Nes 72	Total file size 172.6 GB	Select	:
DB1010000000000000000000000000000000000	0	File name 🔿	Dete type	File size	Last modified	File evaluability	
•         March March March March March March March         400 March Mar	0	5259212998027AQAGCTCTKGAQCTCchaft tar	Filtered Raw Data	153.8 MR	6 OCT 2021	O Archived	:
	0	5259213040027AGAGCTCTbGAGCTC.textp.gr	Raw Data	22.7 68	2 007 2021	Archived	:
•         Million 2008/MCRC 1104/2007/Mate         Examples data         014         6.072.00         c. Annuel III.           •         Million 2008/MCRC 1104/2007/Mate         Readmark for Annuel III.         All All Biological III.         Annuel III.         III.           •         Million 2008/MCRC 1104/2007/Mate         Readmark for Annuel III.         All All Biological III.         Annuel III.         III.           •         Million 2008/MCRC 1104/2007/Mate         Readmark for Annuel III.         Million 2008/MCRC 1104/2007/Mate         III.         IIIII.         III.         III.         III.         III.         III.         III.         IIII.         IIII.         IIII.         IIII.         IIII.         IIII.         IIII.         IIIIIII.         IIIIIIIII.         IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	0	5259212990027AGAGCTCTxGAGCTC.fiber_ond- WETAGENOME.sh	Filtered Ron Gola	6.48	6 007 2021	C Archived	:
International Control (International International Internationa	Ö	525921.3098027AGAGCTC-TAGAGCTC-Ilter- WETAGENOME.faste.gz	Fibered Raw Data	17 GB	6 007 2021	O Archived	:
International Control Contente Control Control Control Control Control Control	0	5259213910027AGAGCTC-TAGAGCTC-fibered-methods.txt	Filtered Rev Osta	1.4 82	6 007 2021	- Archivad	:
Image: 2008/0000000000000000000000000000000000	٥	\$259213898027AGADCTCTbDAGCTC58ared report to:	Filtered Row Data	3648	6 007 2021	G Archived	:
Image: Description of the state of	0	Table, 0,-, 2000040504.taxanomic, composition.tat	Metagenome Report Tables	16.5 KB	16 OCT 2021	O Archived	:
V         Multiple (March 1000)         Multiple (March 10000)         Multiple (Marc	0	Table, 9 3300048584 functional, diversity, bit	Metagenome Report Tables	31.8	18 OCT 2021	O Archived	:
Management data feasibility         Name of the Mark 200         Na		Melogeneme Dath Assembly ZJ_Prado_148		Number of files 72	Tetal file size 216.1 GB	Select	:
V bit service         Notified and the s		Metagenome-Draft Assembly ZJ_Prado_202		Number of files 72	Tetal file size 163.3 GB	Select	:
V Magenese-Staf Ausseldy Subsection 4 Text Res 2014 Star Star Star Star Star Star Star Star		Metaganome-Draft Assembly ZJ_Prado_199		Number of Nes 72	Tetal file size 106 GB	Select	:
		Metagenere-Daft Assembly ZJ_Prado_4		Number of files 72	Total file size 101.3 GB	Select	:

Materials     Acc by Materials Pro	Explorer						E References	O Documentatio
		si	arch for materials inf	armation by chemistry, co	imposition, or property.			
	M	lateriala o.g. Di-Fe or			0	Seech.		
		10         100           10         100           No         300           No         300           10         100           10         100           10         100	Conversion 1 and 1	Alt Land Elements           bit Land Elements           bit determinis with only these elements           bit determinis<	Keresch         Mark         Mark			
Filters	Reset	All 146,323 ma showing 1-15	terials					Columns *
<ul> <li>Composition</li> </ul>		Natural D	femula	Crystal System	Space Group Symbol	510	Energy Above Hull (cirl(Abov)	Band Gr
Thermodynamics	e:	mp-862630	Az	Horagonal	Plyme	4	0	
Structural Proper	rties	mp-881724	412441	Cubic	Finder	4	0	
Symmetry		np-1183030	ArgAgPb	Cubic	Findan	4	0	
		mp-1182068	AtgC05a	Cuarc	Feder	4	0	
<ul> <li>Calculated Prope</li> </ul>	irbes	ng-183383	Ac <sub>i</sub> Coloe	Dates	Feder	4	0	
<ul> <li>Electronic Struct</li> </ul>	ure	mp-1783088	As_CONg	Cultor	Feder	4	0	
<ul> <li>Magnetism</li> </ul>		mp.882313	AcyOdSn	Cubic	Finder	4	0	
<ul> <li>Elasticity</li> </ul>		10-882786	AujOvDe	Overc	Peder	4	0	
<ul> <li>Surfaces</li> </ul>		mp-861083	AlgOall	Outrie	Finder	4	0	
Disconlasteis		mp-1163120	ArgCuRu	Cubic	Feder	4	0	
· FREDERCOR		mp-667122	AnjCuG	Cubic	Finder	4	0	
<ul> <li>Dielectric</li> </ul>		mp-865470	AtgCu5n	Outric	Findin	4	0	
		10-882583	AtgO#Cu	Cubic	Feder	94	0	
		mp-1183054	Ac <sub>i</sub> GeS	Cubic	Findan	4	0	

### Generalized Data Portal Layout

## **Towards Generalized UI Flow Patterns**





#### 

### Generalized Optimization UI Flow



## Scientific Software Dimensions

Subset of the dimensions and projects we are exploring

Project	Users	Domain	Main Scenarios	Software Lifespan	Software Team	Technical Stack
DARK ENERGY SPECTROSCOPIC INSTRUMENT	< 1000, internal collaborators	astrophysics	real time data taking	finite, the length of the survey	domain science developers (staff, postdocs, students)	b\$keh
The Materials Project	>200,000, the general material science community	material science	exploring material info	ongoing	domain science developers (staff, postdocs, students)	iii plotly Dash

The decision to use "low code" technical stacks for both projects was based on the software team makeup

## **Towards A Design System Implementation**

## Investigate composing layouts and UI flows in python with a "low code" implementation of the design system.

```
Ð
# Import packages
from dash import Dash, html, dash_table, dcc, callback, Output, Input
import pandas as pd
import plotly.express as px
# Incorporate data
df = pd.read_csv('https://raw.githubusercontent.com/plotly/datasets/master/gapminder2007.csv')
# Initialize the app
app = Dash(__name__)
# App layout
app.layout = html.Div([
    html.Div(children='My First App with Data, Graph, and Controls'),
   html.Hr(),
    dcc.RadioItems(options=['pop', 'lifeExp', 'gdpPercap'], value='lifeExp', id='my-final-radio-item-example
    dash_table.DataTable(data=df.to_dict('records'), page_size=6),
    dcc.Graph(figure={}, id='my-final-graph-example')
])
# Add controls to build the interaction
@callback(
    Output(component_id='my-final-graph-example', component_property='figure'),
    Input(component_id='my-final-radio-item-example', component_property='value')
def update_graph(col_chosen):
    fig = px.histogram(df, x='continent', y=col_chosen, histfunc='avg')
   return fig
# Run the app
if __name__ == '__main__':
    app.run_server(debug=True)
```

٩	

This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research (ASCR).

Thanks to Deb Agarwal, Lavanya Ramakrishnan, Dan Gunter, Drew Paine, Stephen Bailey, Rollin Thomas, Cecilia Aragon, Devarshi Ghoshal, Ludovico Bianchi, Nan-Chen Chen, Rajshree Deshmukh, Cody O'Donnell

Questions? Contact me at sspoon@lbl.gov



