

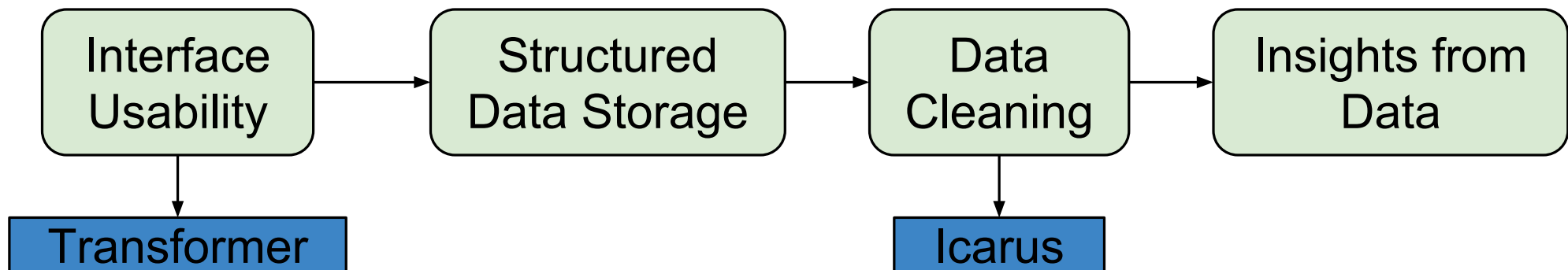
Enabling Effective Data Interaction for Domain Experts

Protiva Rahman, Courtney Hebert, Arnab Nandi

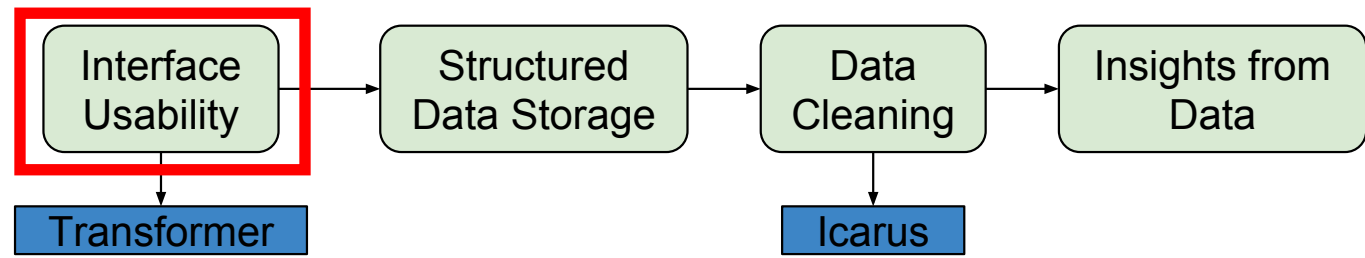
The Ohio State University

Introduction

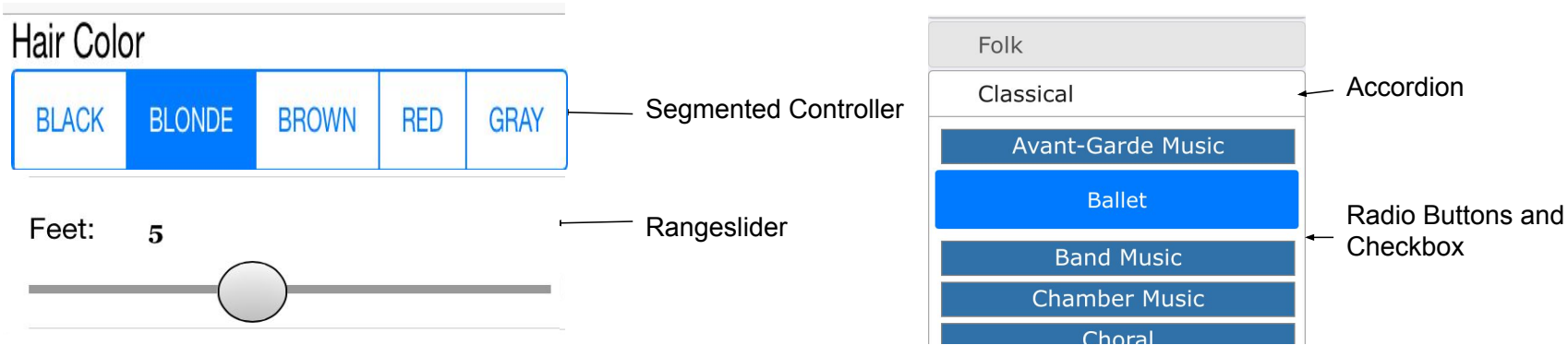
- Increase in EHR + Increase in constrained screen size devices = cumbersome data entry
- Free text annotations next form fields or information in “other” field-> unstructured information -> not available at query time
- More data available for analysis
 - Domain expert interaction needed for cleaning and analysis
 - Domain experts include healthcare practitioners who are not data scientists
 - Need for systems that enable experts to manipulate and gain insights from data



Transformer



- Difficult to fill digital forms on tablets, smartphones and smartwatches
 - Digital forms copied from their paper counterparts -> text heavy
 - Small screen
 - Lack of physical keyboard
- Easier to tap and slide - > use non-text widgets



Transformer (contd.)

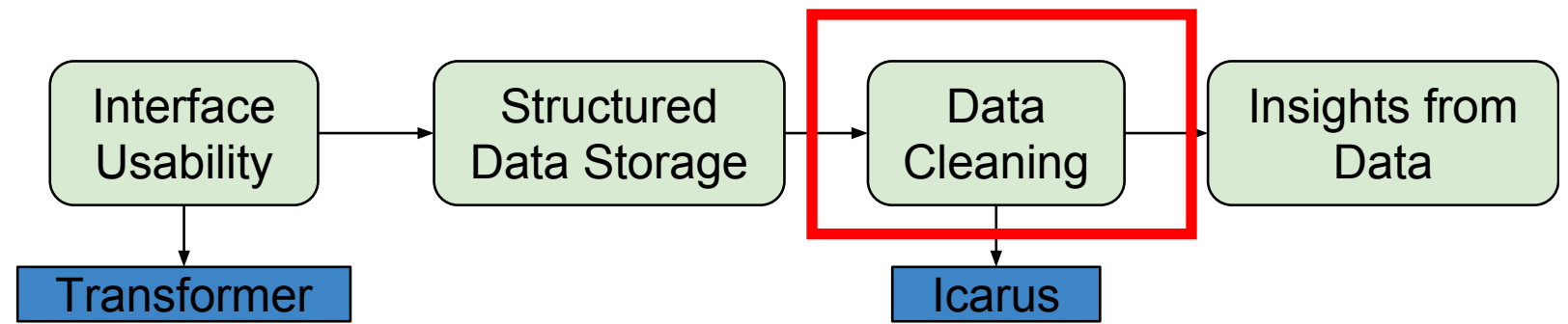
- Use data-driven approach to automatically optimize forms for given screen size
- Forms are an interface to databases -> use data distribution to estimate cost
- Estimate cost of human input

$$cost_i = tap_i \cdot w_{tap} + slide_i \cdot w_{slide} + type_i \cdot w_{type} + scroll_i \cdot w_{scroll}$$

Dropdown	$count \times w_{scroll} + 2 \times w_{tap}$
Toggle	w_{tap}
Text	$avg\ length \times w_{type}$

- Greedy algorithm to minimize overall costs
- 27% decrease in input time with redesigned forms – 15 users, 10 forms

Icarus



- Problem:
 - Microbiology lab reports: Organism sensitive or resistant to antibiotic
 - Unreported sensitivities due to characteristics of organism, antibiotic, and institutional preference
- Automated methods fail
- Domain experts needed to complete dataset via rules
- Large dataset 10,000 rows x 50 columns - Unclear which rules are impactful
- Solution – show informative subsets
- Use database structure to generalize edits to rules

Icarus Interface

ICARUS - Welcome, admin. Undo Save Logout

Rules

Potential Recent Filter Sort

B) Suggested Rules

Enterococcus faecalis Susceptible to Ampicillin + sulbactam (1216)	Yes	No	Maybe
Susceptible to Ampicillin implies Susceptible to Ampicillin + sulbactam (916)	Yes	No	Maybe
Gram Positive organisms Susceptible to Ampicillin are also Susceptible to Ampicillin + sulbactam (916)	Yes	No	Maybe
Enterococcus faecalis Susceptible to Ampicillin + sulbactam	Yes	No	Maybe

Summary of remaining missing values aggregated by column. Current missing is total over all columns

Tobramycin	Ceftriaxone	Cefazolin	Cefepime	Ceftazidime
2124	4152	2465	2130	6994
Daptomycin	Vancomycin	Ertapenem	Doripenem	Linezolid
5626	418	2714	3933	5612
Ampicillin	Ampicillin + sulbactam	Piperacillin + tazobactam	Ciprofloxacin	Sulfamethoxazole + trimethoprim
2533	2810	3600	2130	2306

Current Missing =49547

Dataset

8.96% Refresh Sample Columns: 5 10

A) Editable subset of data

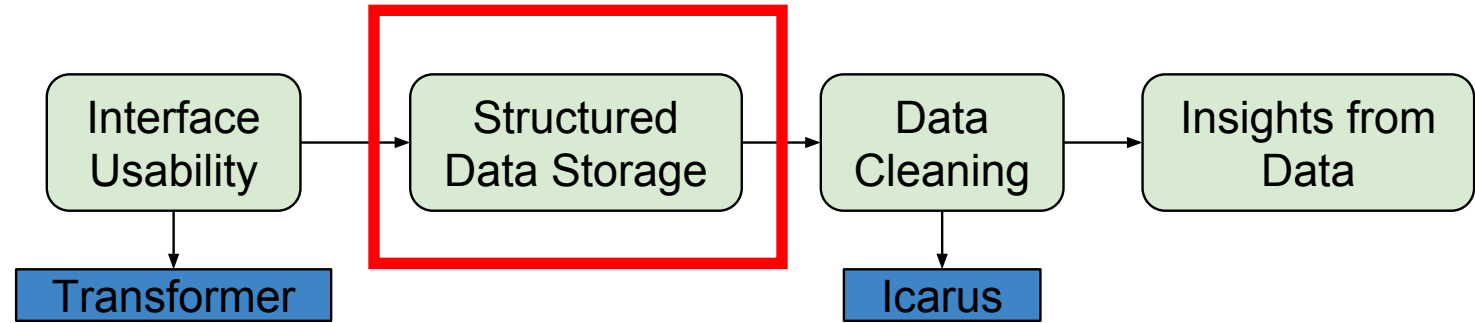
Organism	Cefepime	Tobramycin	Vancomycin	Ampicillin	Ampicillin + sulbactam
Acinetobacter baumannii	S	S	R	R S	S
Enterococcus faecalis	R S	R S	R S	S	S
Enterococcus faecalis	R S	R S	R S	S	R S
Enterococcus faecalis	R S	R S	R	R	R S
Enterococcus faecalis	R S	R S	S	R	R S
Escherichia coli	R	S	R	R	S
Escherichia coli	S	R	R	R	S
Escherichia coli	S	S	R	S	S
Proteus mirabilis	R	S	R	R	R
Pseudomonas aeruginosa	R	R	R	R S	R S

Raw data from microbiology database shown in white background. R- resistant, S-sensitive

Values filled in user rules shown in colored cells

Null Values shown as buttons for the user to fill in with a value

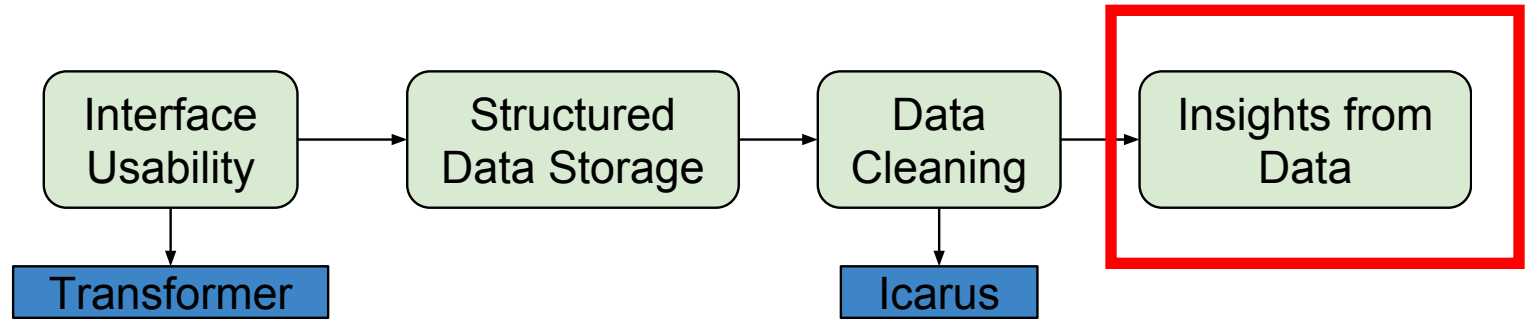
Future Work



Structured Data Storage

- Annotations on form fields are stored as unstructured text and not available for querying
- There are opportunities for automatically structuring annotations using prior data.
- These signals could also be used to update database and form design.

Future Work



Guiding experts to Insights

- Raw data is difficult to digest.
- Exploratory visualizations are iterative and tedious
- What is evident from numbers is not necessarily perceivable on visualizations.
- Find the ideal visualization for a given dataset
- Provide cues to the expert on possible insights

Thank You!