The Use of Collaborative Environments

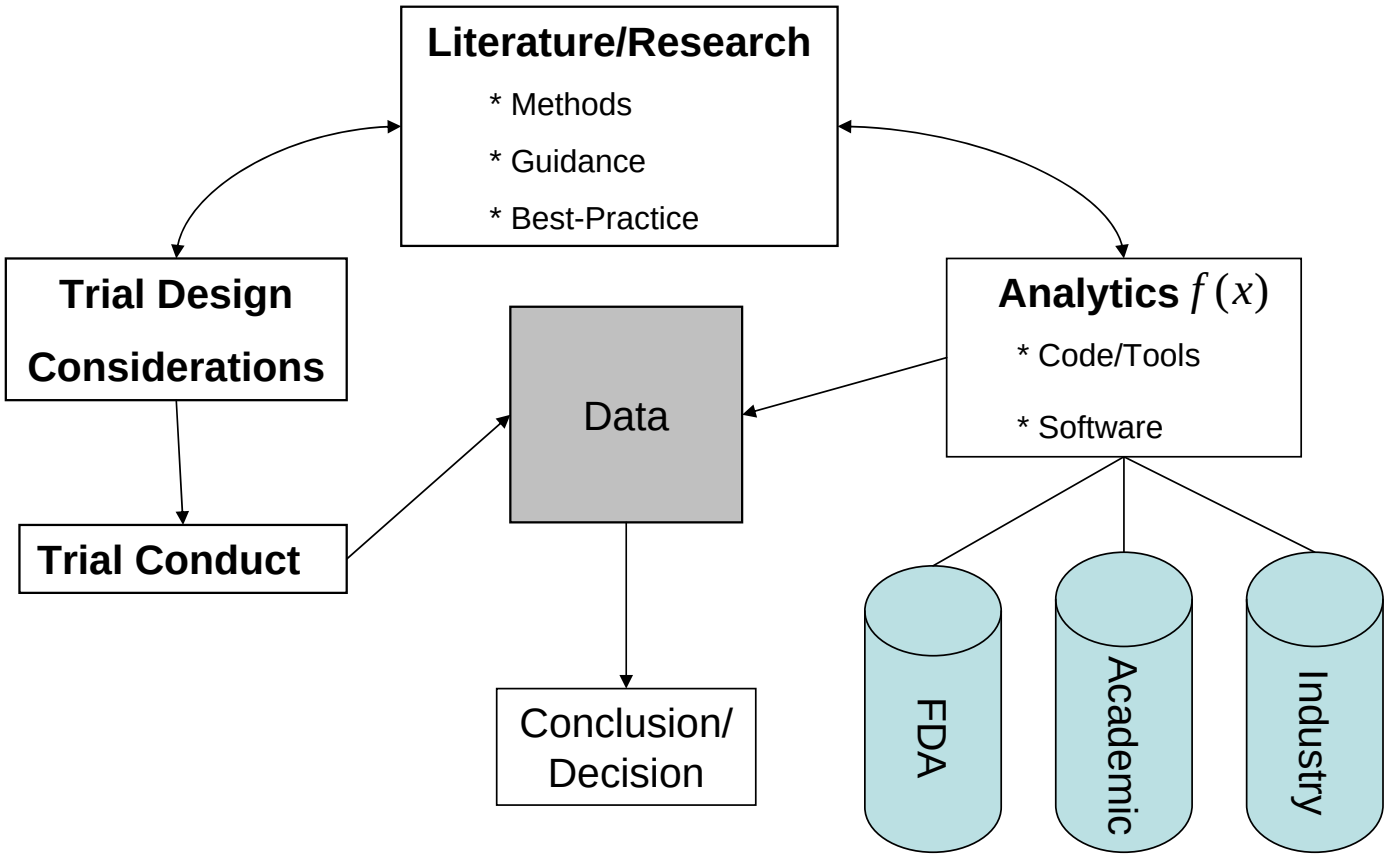
Where We Are Today and Where We Want to Be in the Future

Mat Soukup, Ph.D.
Mat.Soukup@fda.hhs.gov

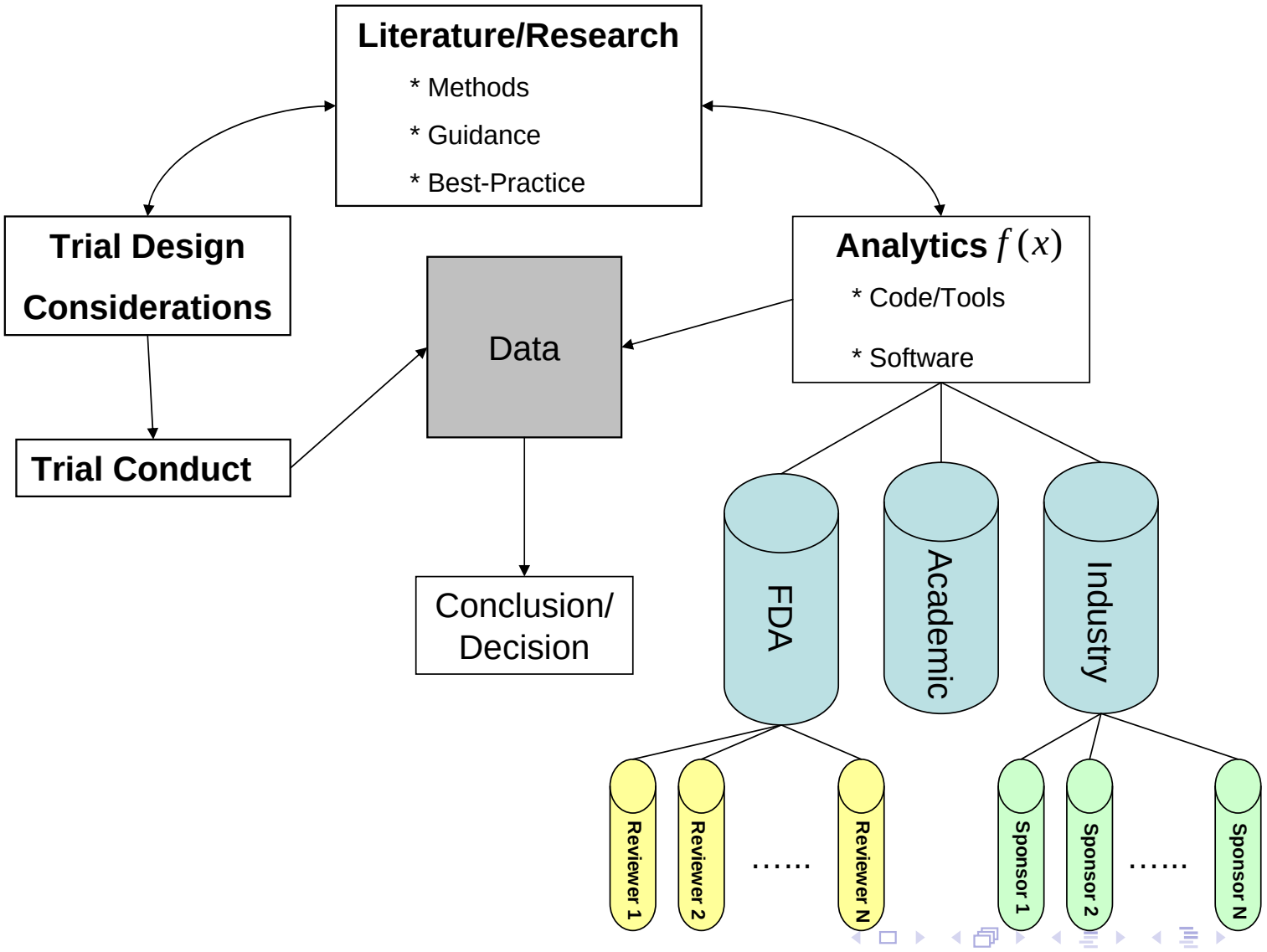
Mathematical Statistician, FDA/CDER

Outline

Defining Today: A Schematic

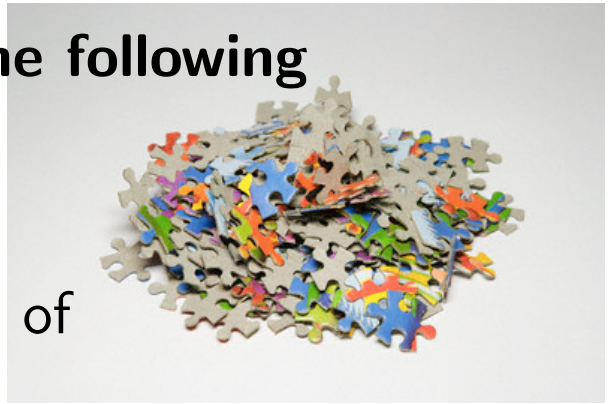


Defining Today: *More Realistic* Schematic



The Problem

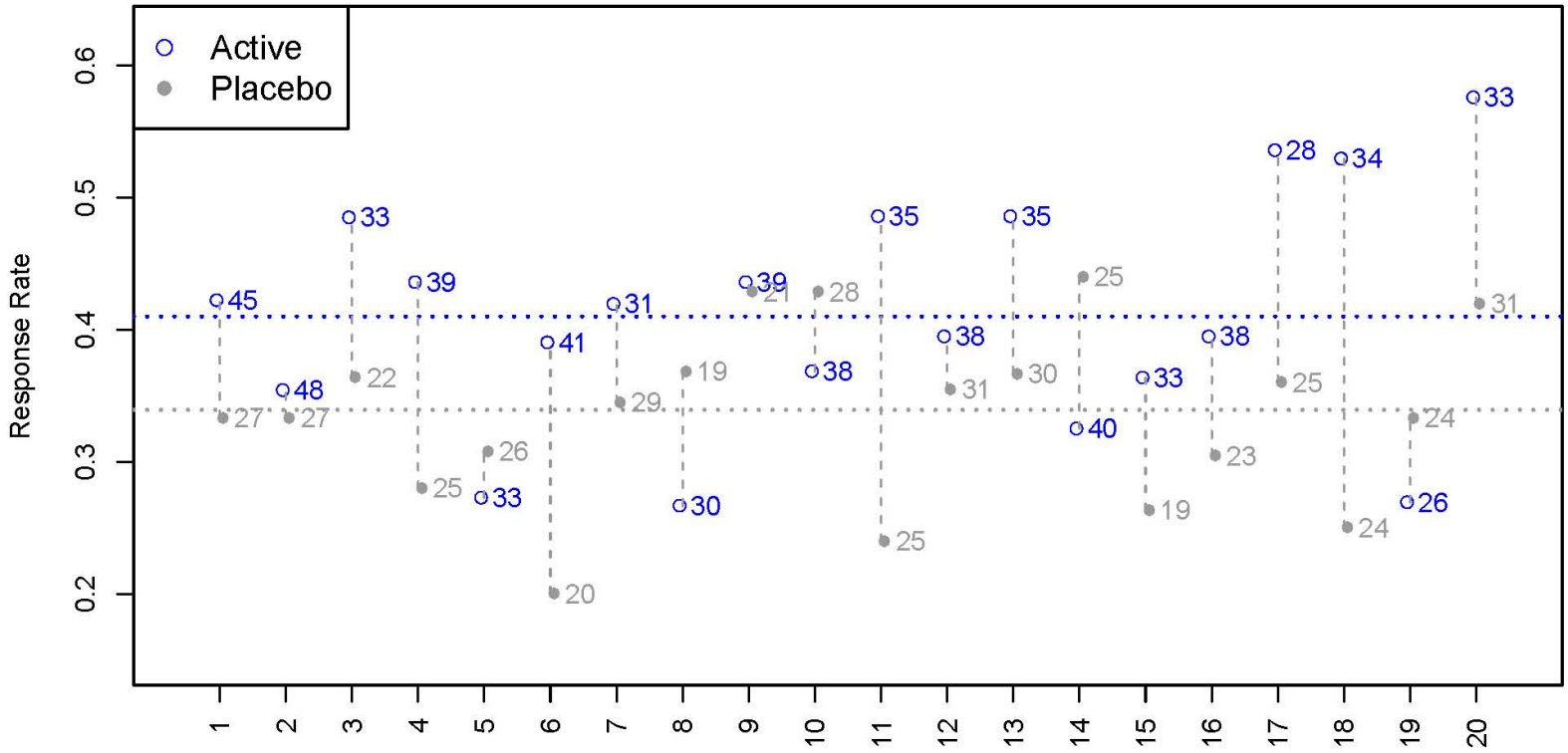
Based upon today's current practice, the following limitations may be present:



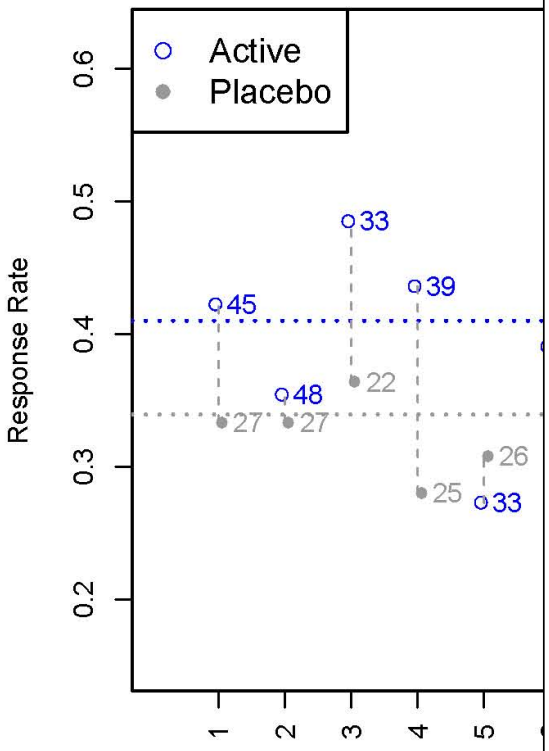
1. Redundancy in analytic development
2. Slow for cross-organization application of literature/guidance/best-practice
3. Quality Control/Validation NOT maximized with limited to no code/open-source software sharing
4. Tendency to rely on antiquated methods
5. Difficult to improve the process for drug/biologic reporting

Illustrative Example: Efficacy By Site

A Graphic...



Illustrative Example: Efficacy By Site



R Code

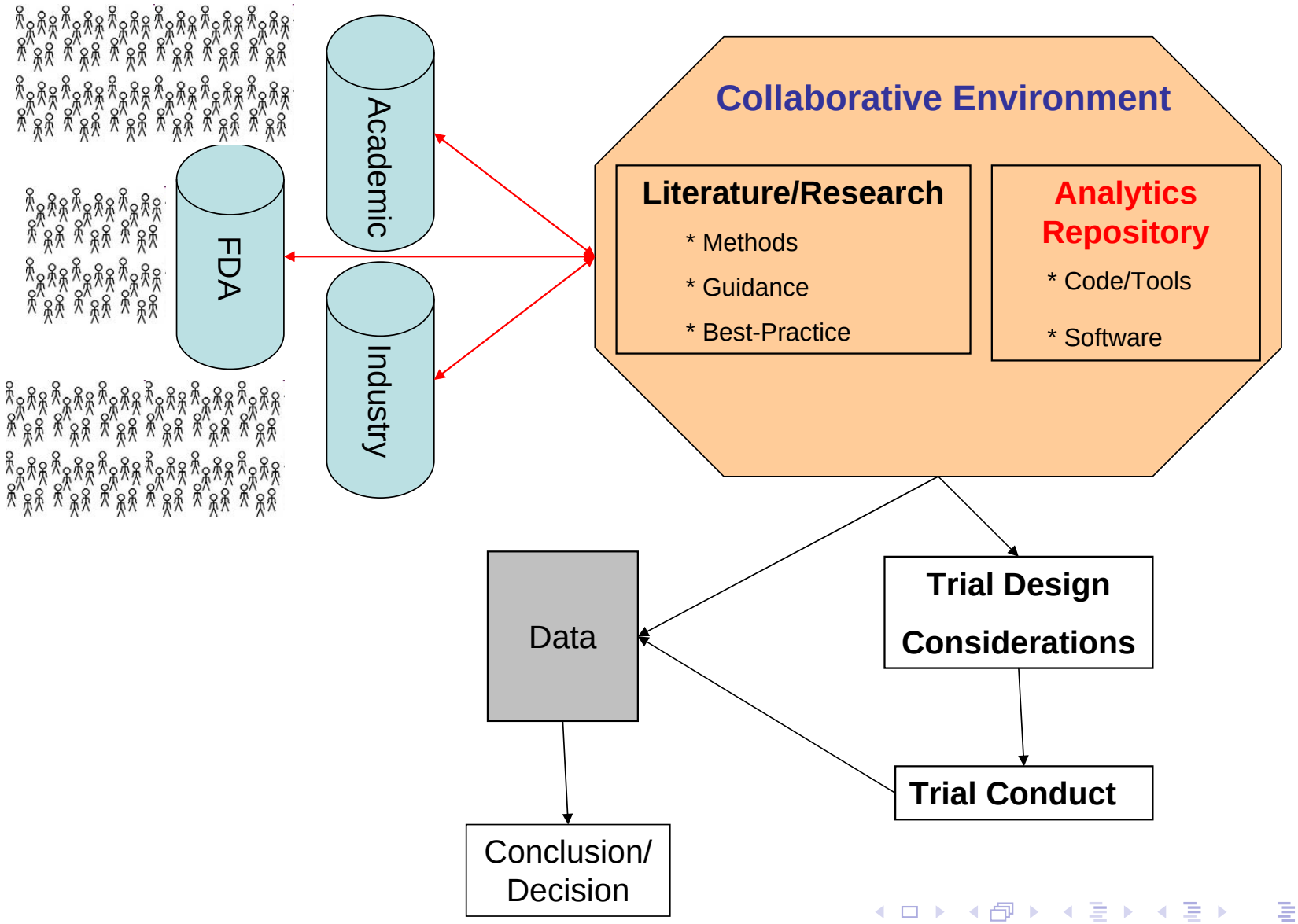
```
"efficacy.by.site" <-
function(yy, site, trt, type="b", legend=FALSE, ...){
    nms <- names(list(...))
    ss <- summarize(yy, llist(site, trt), mean)
    n <- summarize(yy, llist(site, trt), length)

    sdat <- data.frame(ss, n[,3])
        names(sdat) <- c("Site", "Trt", "Mean", "N")

if(type=="b"){
    nsn <- length(unique(sdat$Site))
    ut <- unique(sdat$Trt)
    rnx <- tabulate(as.factor(sdat$Site))
    sdat$plotx <- rep(1:nsn, rnx[rnx>0])
    # Creation of the figure.
    if("ylab" %in% nms)
    plot(c(.5, nsn+.5), c(min(sdat$Mean)-.05, max(sdat$Mean)+.05),
            type="n", axes=FALSE, ...)
    else plot(c(.5, nsn+.5), c(min(sdat$Mean)-.05, max(sdat$Mean)+.05),
            type="n", ylab=paste(deparse(substitute(yy))), axes=FALSE)
    axis(1, at=1:nsn, labels=as.character(unique(sdat$Site)), cex.axis=.75, las=3)
    axis(2)
    box()
    if(length(ut)==2) sdat$plotx <- sdat$plotx + rep(c(-.05,.05), length(sdat[,1])/2)
    if(length(ut)==3) sdat$plotx <- sdat$plotx + rep(c(-.1,0,.1), length(sdat[,1])/3)
    if(length(ut)==4) sdat$plotx <- sdat$plotx + rep(c(-.15,-.05,.05,.15), length(sdat[,1])/4)
    for(k in 1:length(ut)){
        subdat <- subset(sdat, sdat$Trt==ut[k])
        points(subdat$plotx, subdat$Mean,
               pch=trellis.par.get("superpose.symbol")$pch[k],
               col=trellis.par.get("superpose.symbol")$col[k])
        for(j in 1:length(subdat$N)){
                text(subdat$plotx[j]+.3, subdat$Mean[j], labels=subdat$N[j],
                     col=trellis.par.get("superpose.symbol")$col[k], cex=.7)
        }
    }
    for(i in 1:nsn){
        subdat <- subset(sdat, sdat$Site==unique(sdat$Site)[i])
        lines(c(i,i), c(min(subdat$Mean), max(subdat$Mean)), lty=2, col='gray60')
    }
}
if(type=="nonly"){
    nsn <- length(unique(sdat$Site))
    ut <- unique(sdat$Trt)
    rnx <- tabulate(as.factor(sdat$Site))
    sdat$plotx <- rep(1:nsn, rnx[rnx>0])
    # Creation of the figure.
    if("ylab" %in% nms)
    plot(c(.5, nsn+.5), c(min(sdat$Mean)-.05, max(sdat$Mean)+.05),
            type="n", axes=FALSE, ...)
    else plot(c(.5, nsn+.5), c(min(sdat$Mean)-.05, max(sdat$Mean)+.05),
            type="n", ylab=paste(deparse(substitute(yy))), axes=FALSE)
```

Illustrative Example: Efficacy By Site

- Is the approach publicly available or does the public know about it?
  - Potentially, it's been presented at several professional meetings.
- How to reproduce this visual representation?
  - Write your own code; ask the author.
- What if there are ways to improve the representations?
  - Publish/present at public meeting
- What if you have written sleek code, can you share it?
  - Not really; potentially with the author
- What if the code is written in a language my closed system does not run?
  - Rewrite it!

Outline

Collaborative Schematic



Rely on a Community *not* Individual Organizations

Advantages

- Transparency
- Increase the talent pool
- Current; documents/materials/code can evolve
- Efficient - evolution towards improvement (not reproduction)
- Addresses needs of participants; tailored towards them

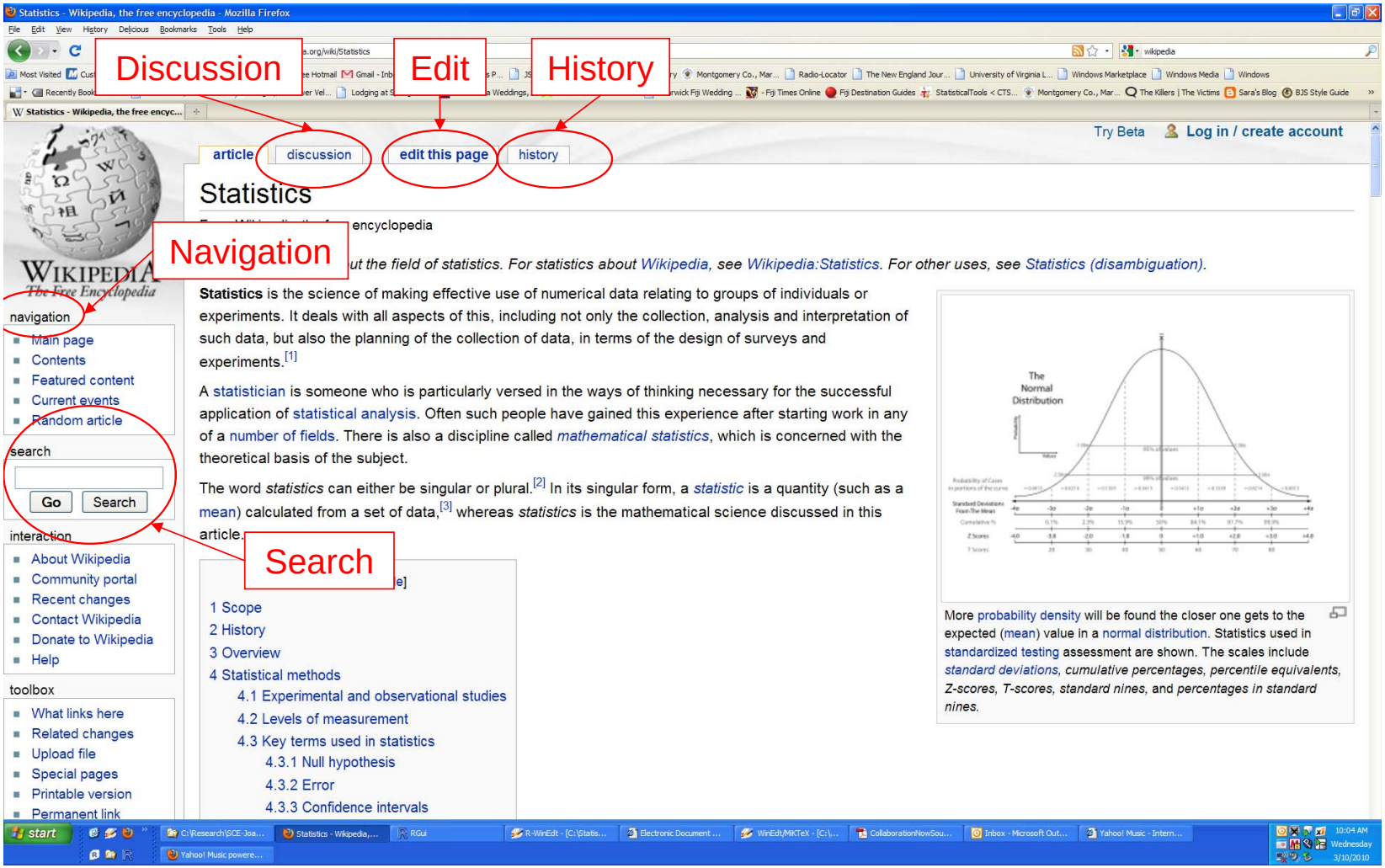
Disadvantages

- Trustworthiness
- Lack of authority
- Content is driven by willingness of the community to share
- Abundance of information; potentially redundancy

Solution: The Wiki Way?

- Most popular and HIGHLY successful Wiki: **Wikipedia**.
- **Definition**: A wiki is a website that uses wiki software, allowing the easy *creation* and *editing* of any number of interlinked Web pages, using a simplified markup language [source: Wikipedia].
- Creation/Editing is done via the web browser - no fancy software is required.
- **Community** of users add/edit content $\longrightarrow$ pages/website is not static but **ALIVE**!
- *Invokes* user participation to create or collaborate.
- Subject to GNU-GPL regulations making them free software programs.

Wikipedia Screenshot

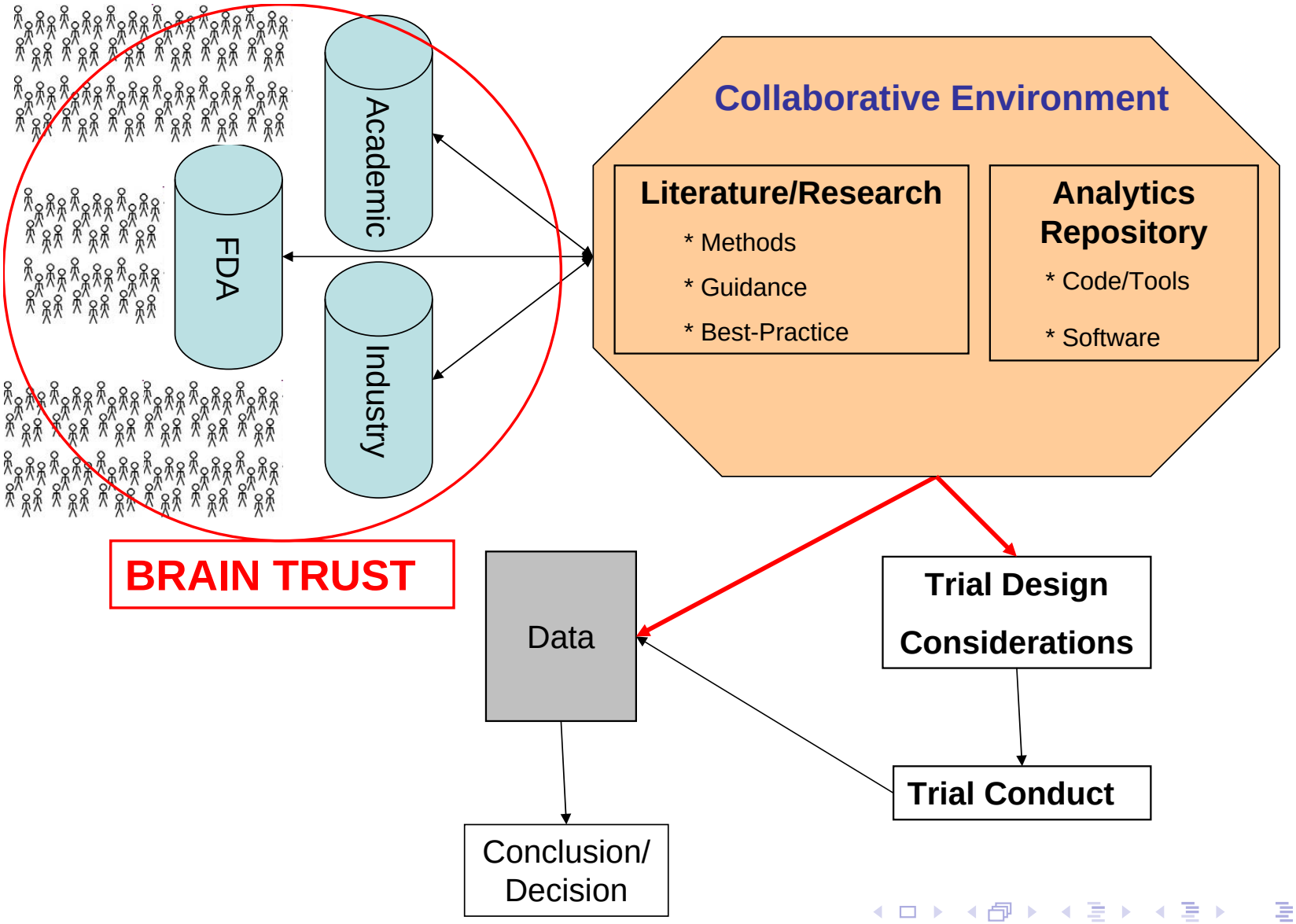


Wiki: Strengths and Weaknesses

- Bad content may appear from time to time
  - 50% of mass deletions were modified in less than 3 minutes (Wikipedia, CHI 2004)
- Lack of contributions to important topic areas
- Topics which are emerging can **evolve quickly**
- **Rewards** contributor to know their efforts are being utilized by others
  - Can improve by rewarding important efforts
- Lack of citation/recognition for wiki contributions
  - Recently; may be better acknowledgements of such contributions
- **Development** in topics otherwise not known to originators

Outline

Bringing Stakeholders Together



General Considerations

In the development of a collaborative environment (wiki) the following considerations should be made.

- Target the consumer - meet the needs of the end-user
- Move from a culture of a lack of information to a culture of information overload - site organization/knowledge management is essential.
- Hierarchy is horizontal - does *not* require authoritative approval for participating (may be some guidelines within institutions)
- General acceptance that content does not have to be "finished" to be published
- Training of wiki use may be required
- Requires buy in from upper management

Utilizing CTSpedia

- Building on the work of the CTSA, advance CTSpedia for a broader audience
- Faces several challenges
  1. Site needs to meet the needs of all users
  2. Limited Resources: financial and human - will need to increase as content grows to ensure organization of site maintains its intended functionality
  3. Reorganization
  4. Employees within an organization should be encouraged to freely contribute - potential culture change
- May need to limit the scope - to start! Keep initial development towards targeted needs

Partnering CTSpedia and FDA

- Computational Science Center (CSC) Project Team 17 proposed to "Establish a **collaborative environment** for development of specialized program code (like SAS or SPLUS scripts or compiled code) for analytical tools."
- CTSpedia approached by FDA CSC Project Team 17 in September 2009
- Since January of 2010, CTSpedia and FDA core members have been participating in restructuring CTSpedia.
- Several challenges need or needed to be addressed
  1. Reorganization of an active wiki towards one planned for a wider audience
  2. Legal considerations (FDA)
  3. Increasing resources available to the project
- **Continues to move forward**

What the Future *Might* Look Like

Repository Title Name

Institution

Jump | Search

Register | Log In

Home Page
Tools
Graphics
Best Practice
Development
Other Links

Wiki Links
Sandbox
Frequently Asked Questions
Others

Edit | Attach

Analytical Tools
Tool Repository
Recently Added Tools
Most Used Tools
Highest Rated Tools

Tool Development
Tool Needs
Tools Being Developed
Tools Completed

Graphics
Graphic Tools
Recently Added
Highest Rated

Best Practices
Listing
Recently Added

Other Links
Announcements
Others…

Edit | Attach | Print Version | History: … | Backlinks | Raw View | Raw Edit | More topic actions

How a Tool Page *Might* Look

GeniusMIApproach

Description

This tool allows one to implement the multiple imputation procedure developed by Billy published in Journal for Cool Statistics. The strength of this approach is for count data collected over a series of visits.

Link to article

Code Information

Program Language: R version 2.8

Tool Version: Version 1.1

Tool Meta Data: GeniusMIApproach-Meta

Link to Meta Data Page

Tool Example: Coding Examples

Link to Examples in Meta Data Page

Tool Code: GeniusMIApproach

Link to read-only code

Key Words

Multiple Imputation, missing data

Author

Email: Billy@StatsRock.org

Organization: StatsRock, Ltd.

User Reviews

Number of Reviews: 10

Link to reviews

Average Rating: ★★★★☆ (4 of 5)

Discussion **Edit** **History**

Number of page views: 0 0 4 1

Special thanks to the following people

FDA	CTSA
Sue Bell	Mary Banach
Joan Buenconsejo	Laurel Beckett
Chuck Cooper	Frank Harrell
Gary Gensinger	Jeffrey Horner
Ted Guo	Sally Thurston
Don Salzer	
Norman Stockbridge	
Chris Tornoe	

Thank You!