Microsoft Office Documents as R-Applications

Josh van Eikeren and Paul van Eikeren

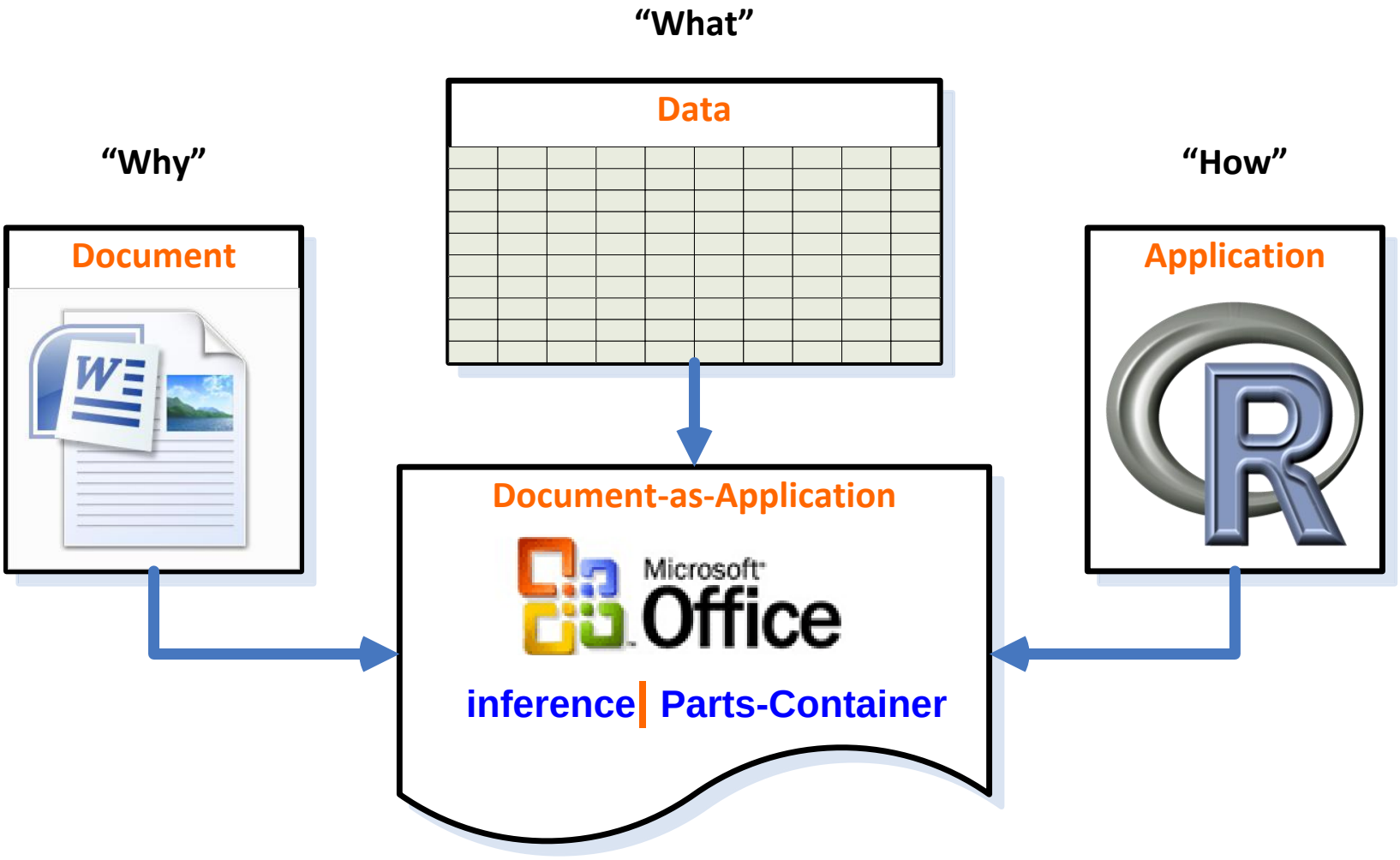
userR! 2009 (July 8-10, 2009)

inference
from Blue Reference

Problem to be Solved

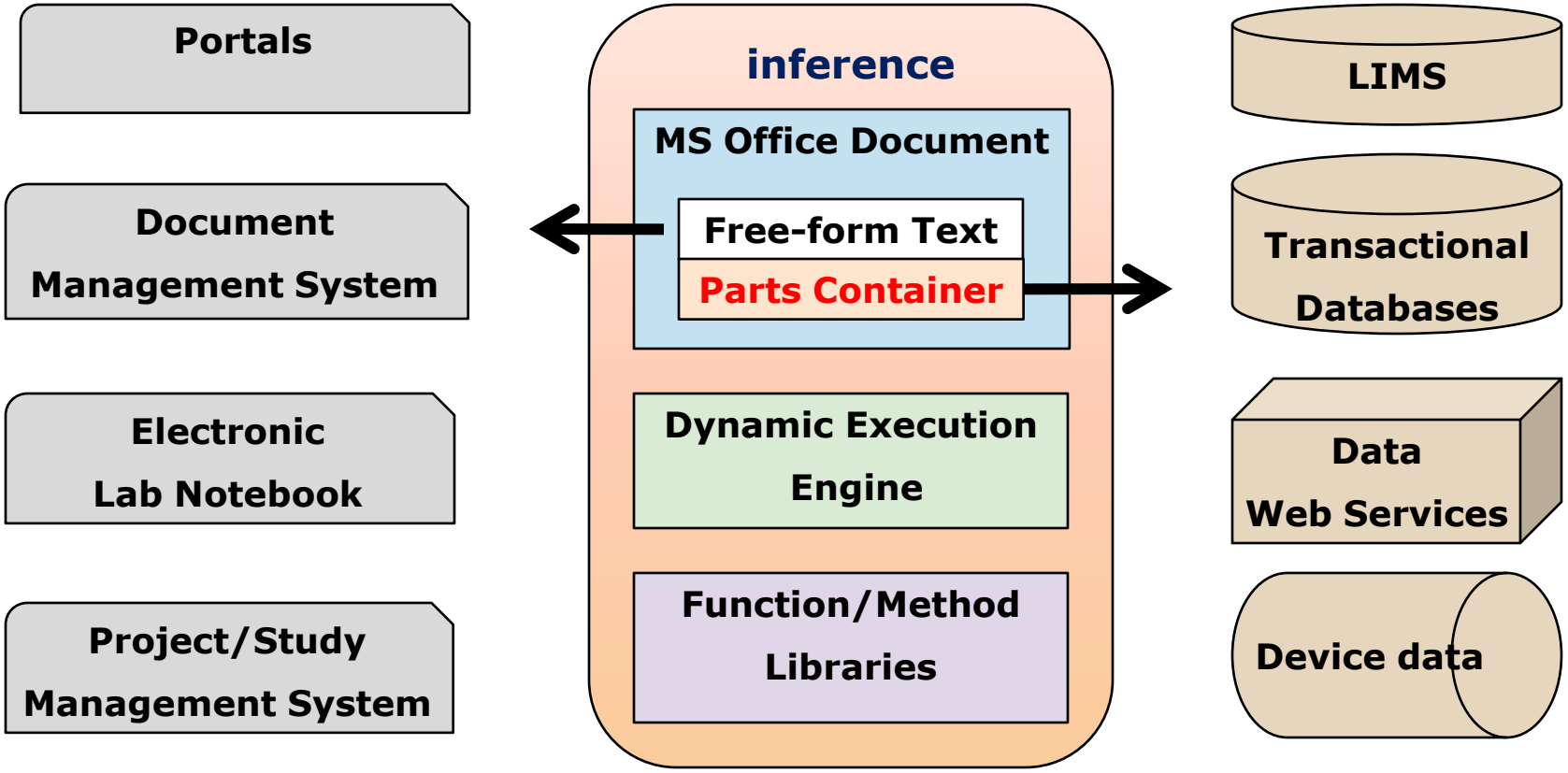
Industry: Reproducible R & D

inference
from Blue Reference



Industry: Bridge Two Worlds

inference
from Blue Reference



Free-form Text

Document-centric

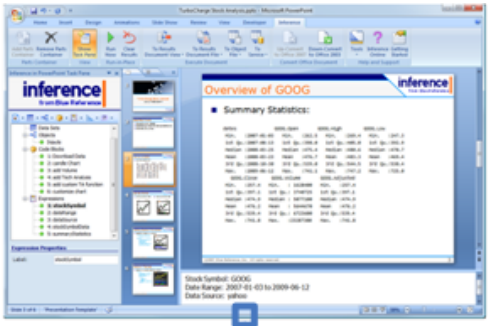
Structured Data

Process-centric

Solution approach

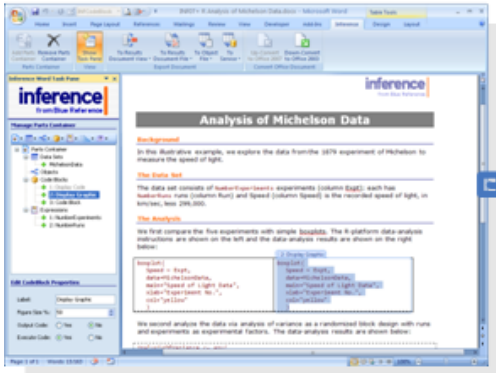
inference: Parts Container

inference | from Blue Reference

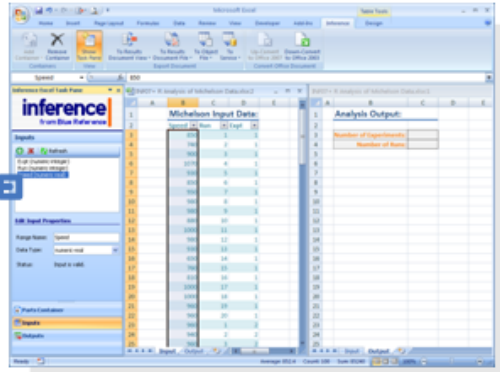


inference | in PowerPoint

inference | in Word



inference | in Excel

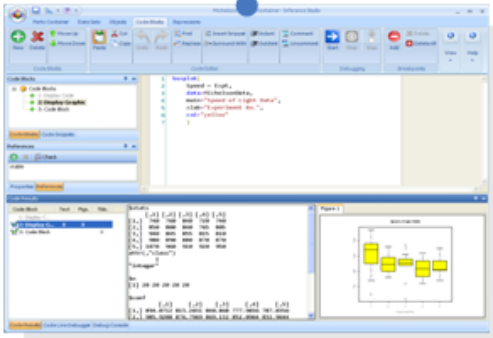


inference | Parts Container

- Data Sets
- Objects
- Code Blocks
- Expressions

R Statistical Environment

R Packages



inference | Studio

inference in Word

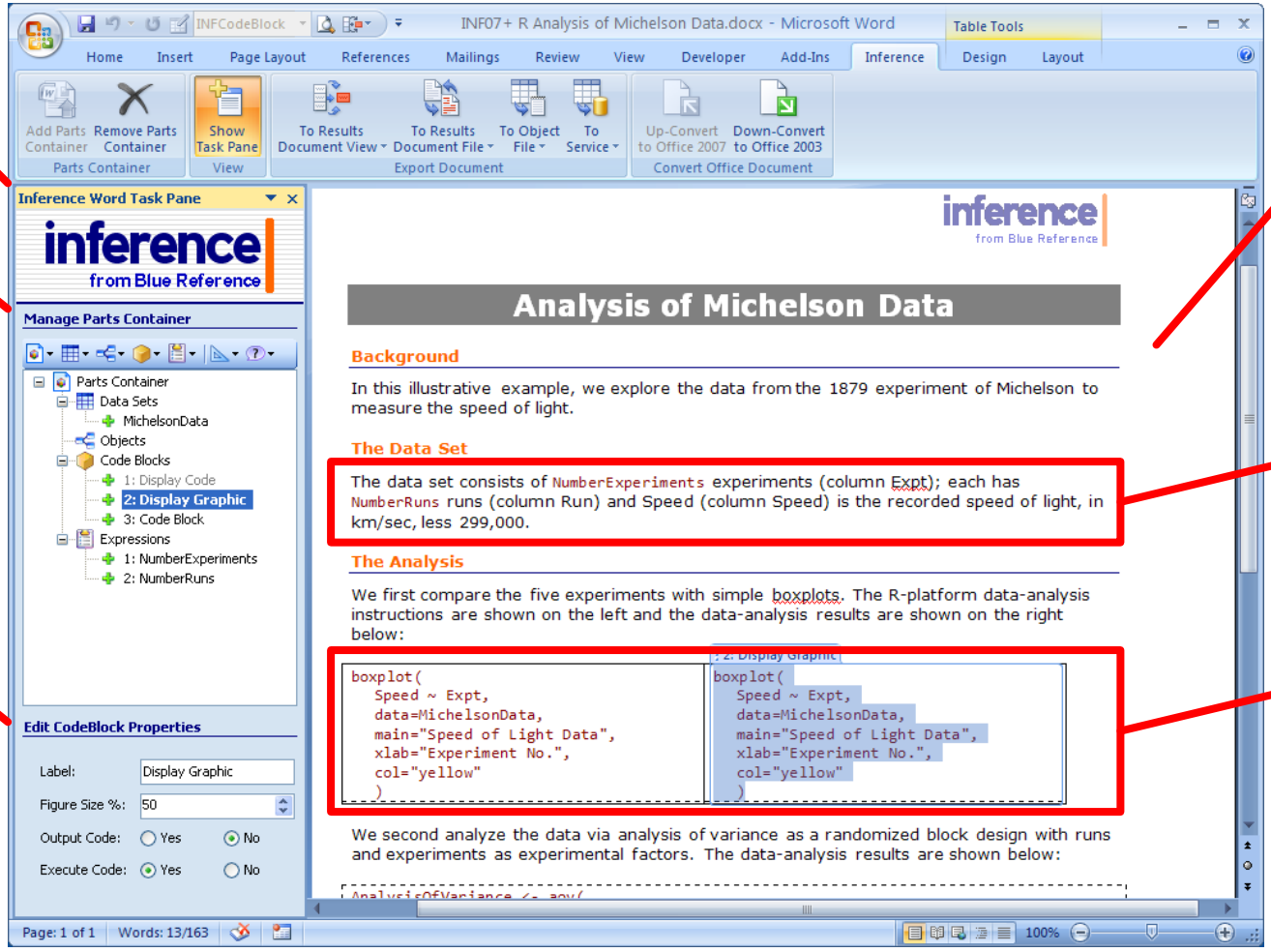
inference in Word: rationale

inference
from Blue Reference

- **Leverage Systems You Already Have**
  - Microsoft Office on every desktop (98% of corporate desktops)
  - 500 Million users of Microsoft Office
- **Focus on Supporting the Work Process**
  - Reusable protocols as “best practices”
- **Documents as Work Process Solutions**
  - Documents at the core of industry workflow
  - Technical workers comfortable with Microsoft Office (Word, Excel and PowerPoint)
  - Documents leverage existing systems (document management, SOPs, regulatory practice)
  - Technical workers still like paper!

inference in Word: elements

inference
from Blue Reference



inference in Word: dynamics

inference from Blue Reference

Source document

inference from Blue Reference

Quiver Plots: Visualization of Vector Fields

The Data Set

Vector fields play an important role in science and engineering. They are used to describe a wide variety of phenomena including fluid flow. To illustrate vector fields, let us consider a collection of x,y,z values that describe a surface net. The z value for each node is calculated from the function

$$f(x,y) = (3x^2 + y) \times e^{-(x^2+y^2)}$$

The net is displayed in three dimenstions in the figure on the right. A vector field would correspond to the magnitude and direction of the slope of the net at each node.

```
displayData()
```

The Analysis

Large vector fields often exhibit quite complex structures, which can be difficult to reveal, making efficient visualization of a vector field an important analysis. A preferred 2-D visualization, illustrated on the right, corresponds to a "quiver" plot. You may recall that a quiver is a carrying case for arrows.

```
2: display quiver plot
quiver2(
  f2,
  x,
  y,
  color.palette=terrain.colors
  )
```

The Interpretation

The analysis calculates a dx and dy component of the slope at each node in the net. It then uses this information to define a vector (an arrow) at each x,y position, whose length and orientation correspond to the magnitude and direction of the slope, respectively

Results document

inference from Blue Reference

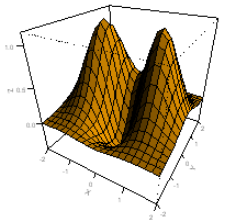
Quiver Plots: Visualization of Vector Fields

The Data Set

Vector fields play an important role in science and engineering. They are used to describe a wide variety of phenomena including fluid flow. To illustrate vector fields, let us consider a collection of x,y,z values that describe a surface net. The z value for each node is calculated from the function

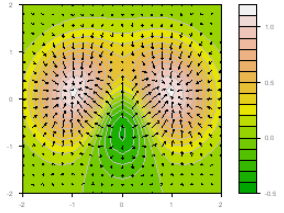
$$f(x,y) = (3x^2 + y) \times e^{-(x^2+y^2)}$$

The net is displayed in three dimenstions in the figure on the right. A vector field would correspond to the magnitude and direction of the slope of the net at each node.



The Analysis

Large vector fields often exhibit quite complex structures, which can be difficult to reveal, making efficient visualization of a vector field an important analysis. A preferred 2-D visualization, illustrated on the right, corresponds to a "quiver" plot. You may recall that a quiver is a carrying case for arrows.



The Interpretation

The analysis calculates a dx and dy component of the slope at each node in the net. It then uses this information to define a vector (an arrow) at each x,y position, whose length and orientation correspond to the magnitude and direction of the slope, respectively

inference in Word: tailored solution

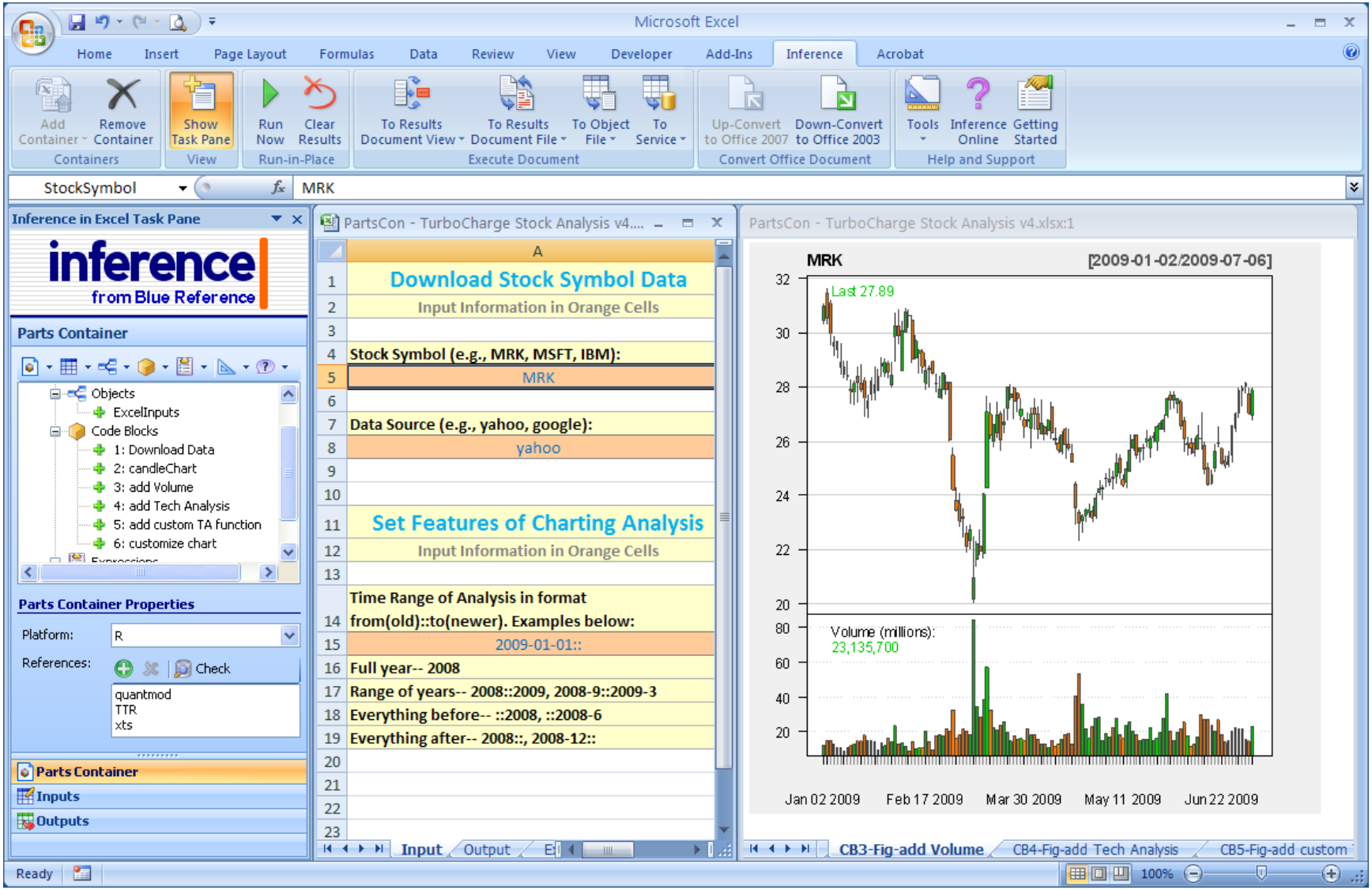
inference
from Blue Reference

- Users only see the capability that they need
- Shallow learning curve using Office
- Data integration and management tightly integrated
- Analysis protocols provide audit trail for how results obtained
- Analysis and documentation occur concurrently
- Automated generation of Results Documents
- Adaptable and extensible using the Inference platform
- Inexpensive to set up and maintain because Office already deployed

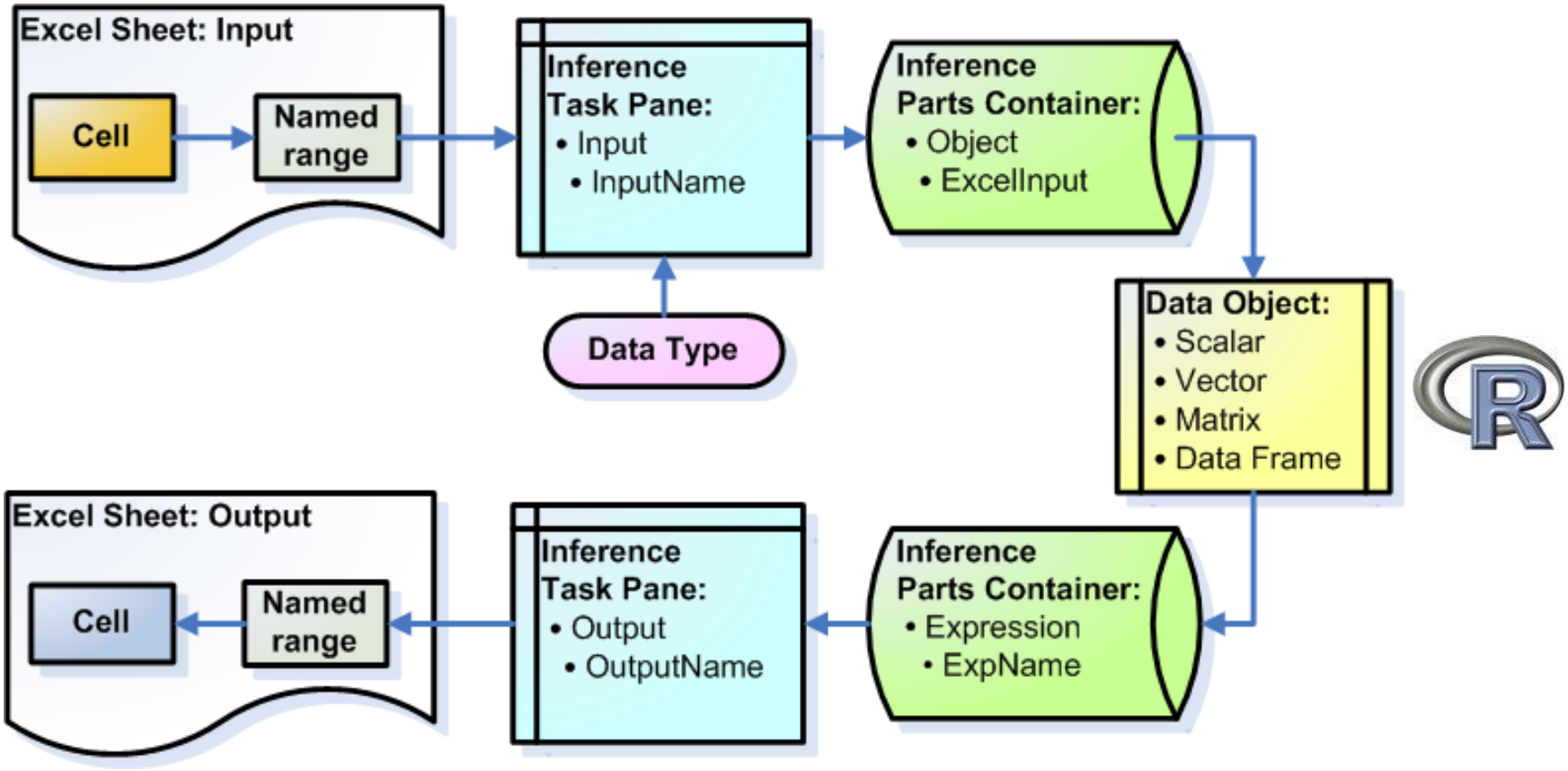
inference in Excel

inference in Excel: elements

inference
from Blue Reference



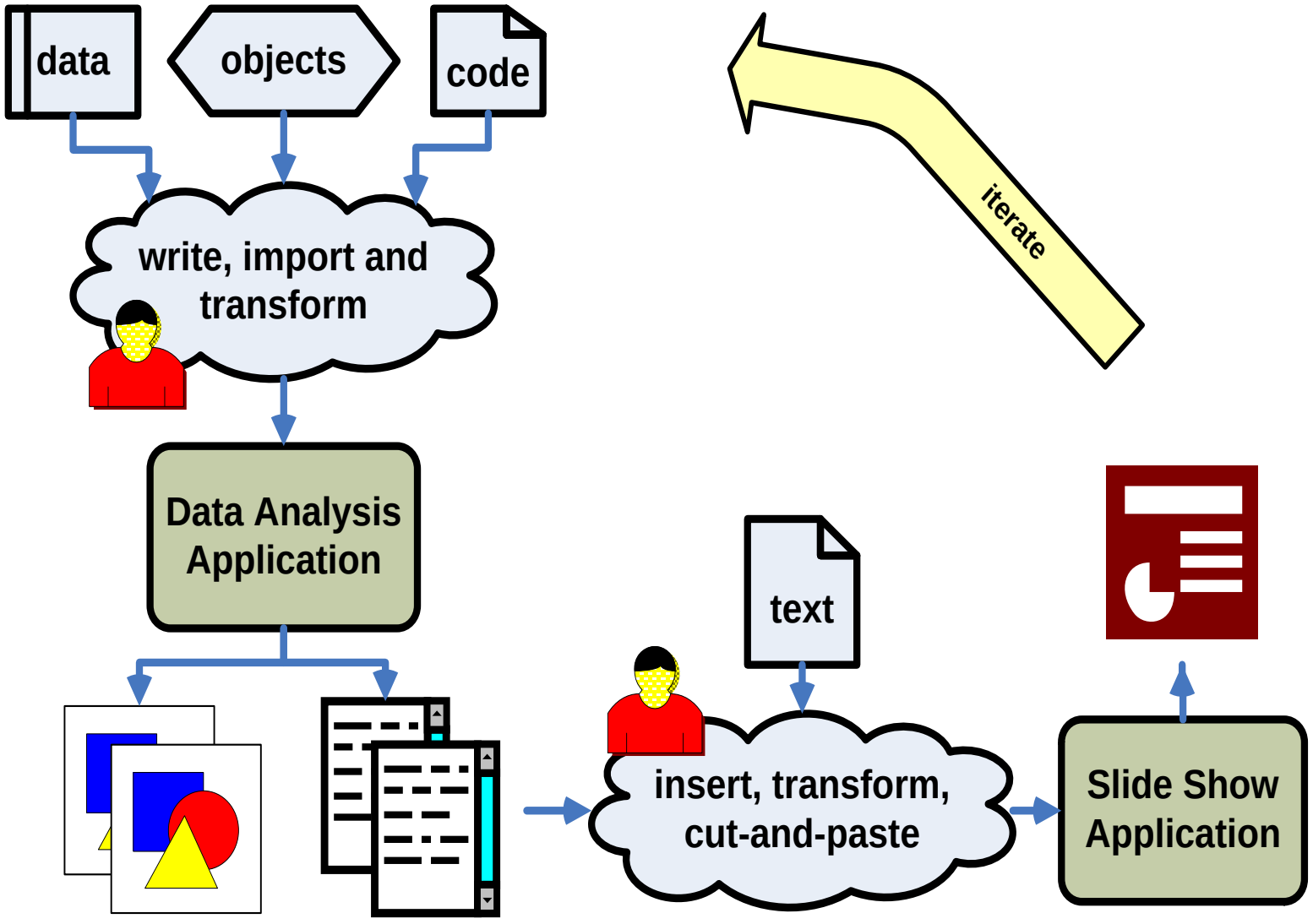
inference in Excel: data flow



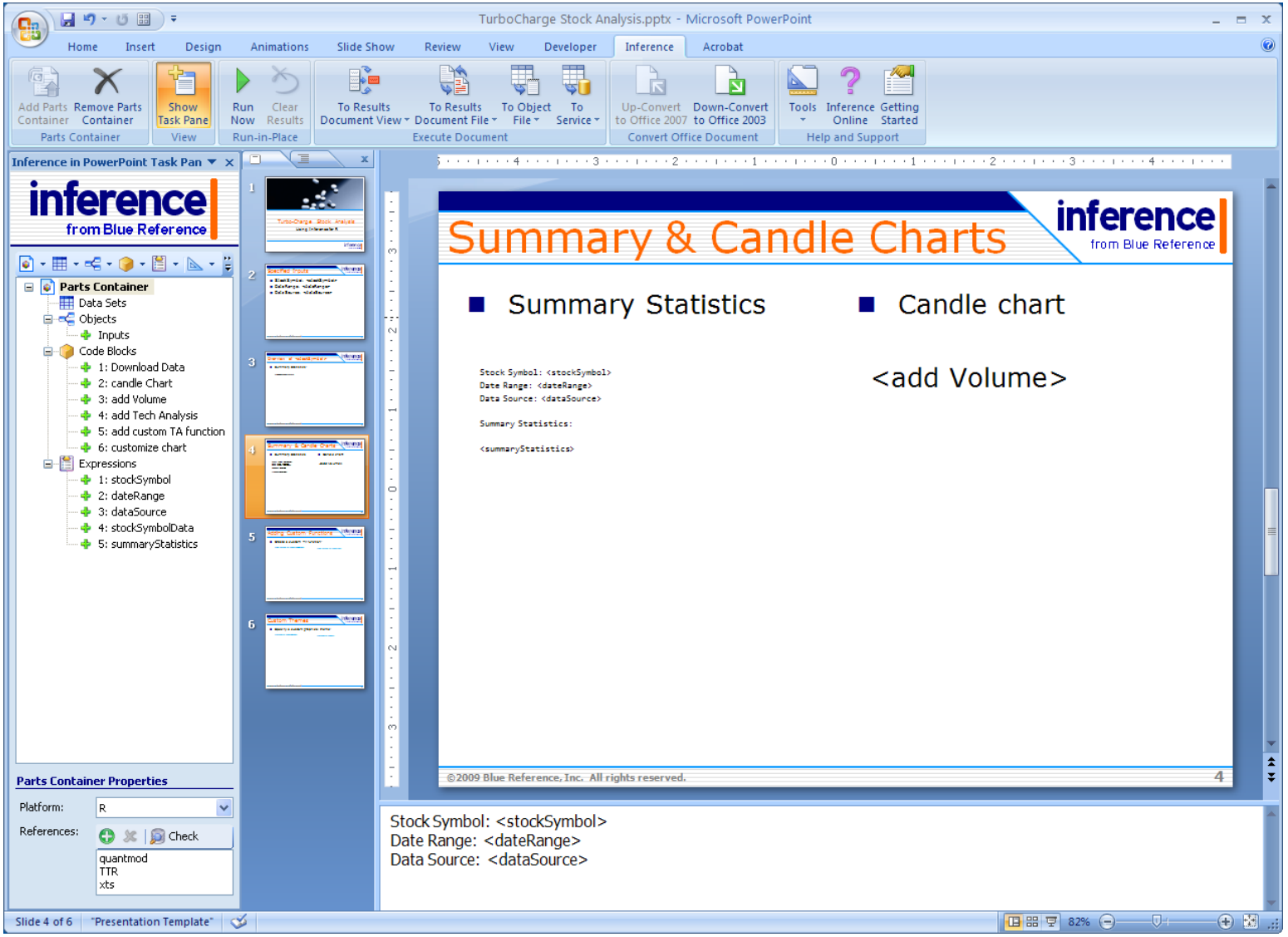
inference in PowerPoint

PowerPoint: conventional approach

inference
from Blue Reference

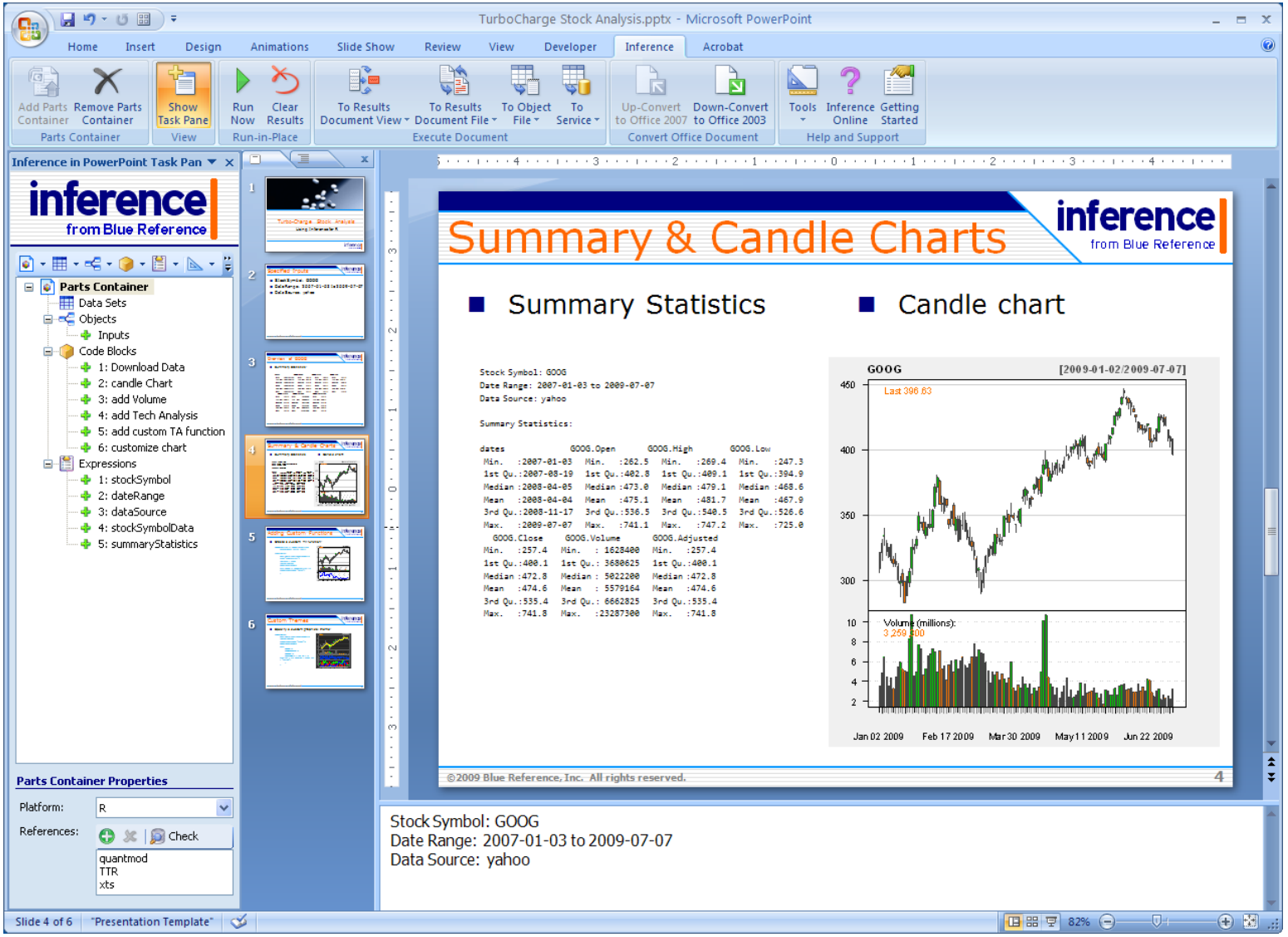


inference in PowerPoint: elements



inference in PowerPoint: results

inference
from Blue Reference



inference in PowerPoint: message

inference
from Blue Reference

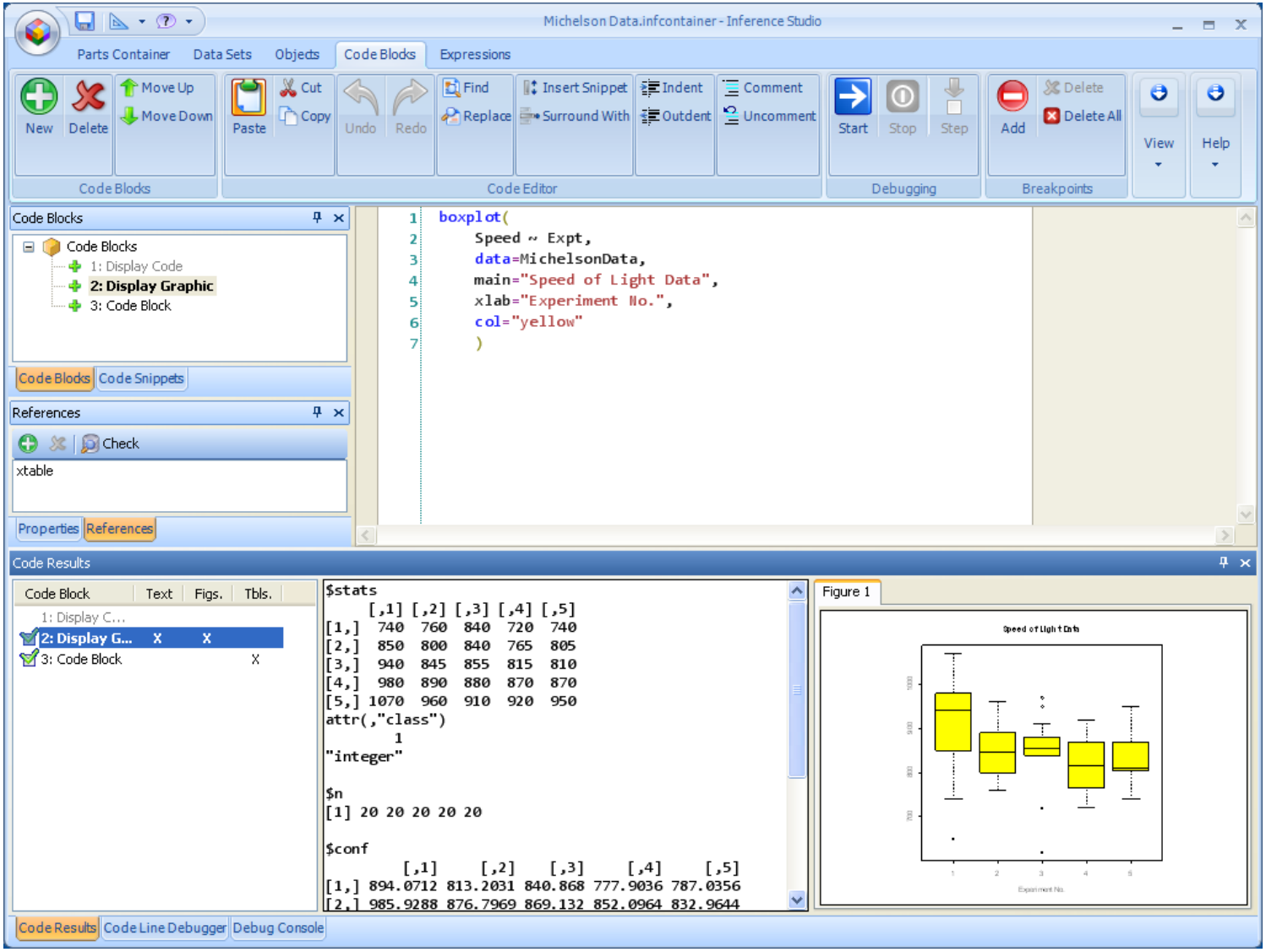
- Lasting presentations
  - all information in one place
  - accessible raw data
  - symbolic and executable solution
  - automated results
  - reproducible and reusable
- Unified and integrated assembly process
  - better
  - faster
  - cheaper
- Presentation with lasting impact
  - no “walled gardens”
  - enables practice of “don’t tell me, show me”
  - learn from “view source” concept

Integrated Development Environment

inference Studio

inference Studio: elements

inference
from Blue Reference

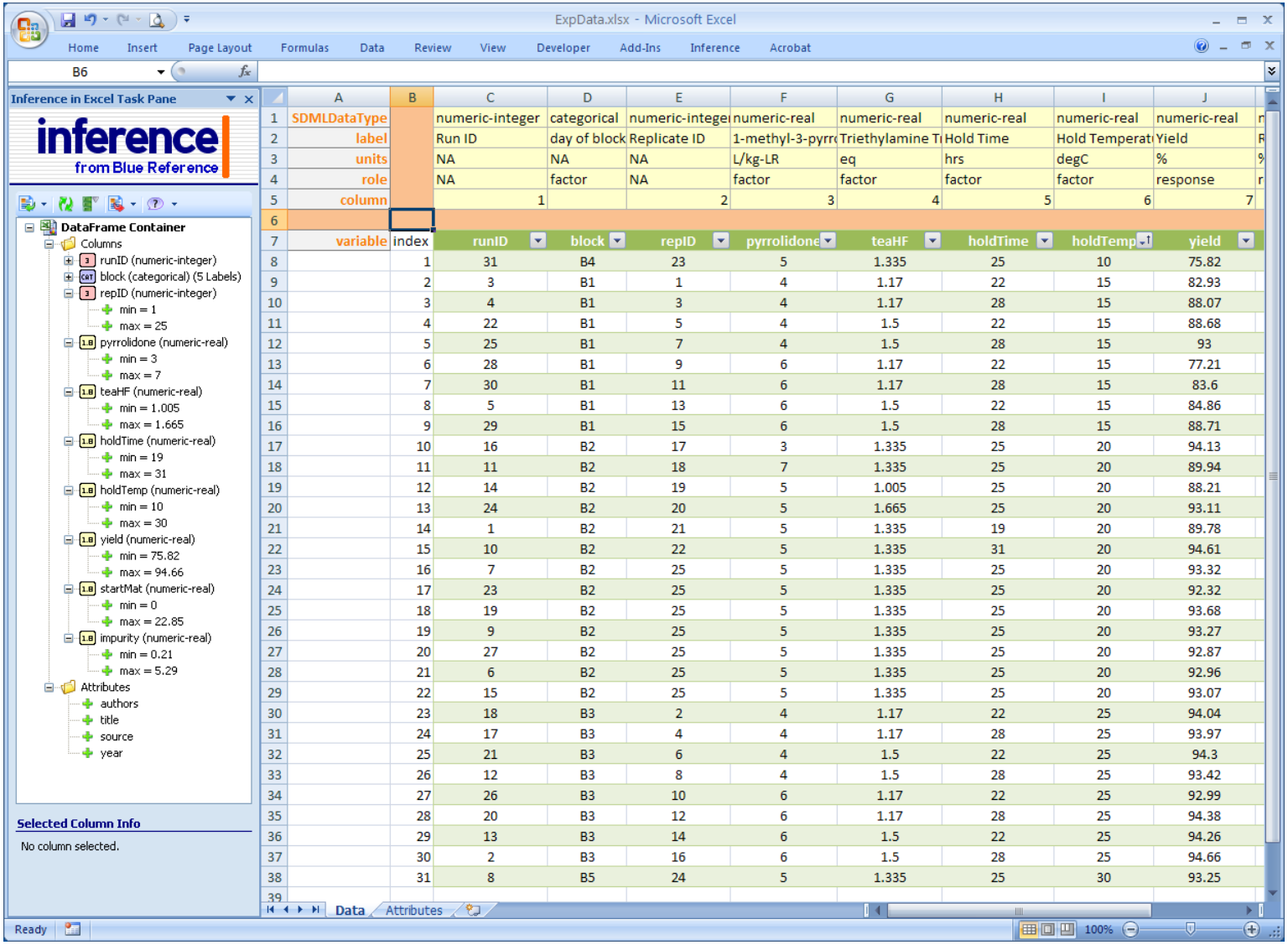


- Excel can be used to manage data frames
- Data frame object: R, .NET and SDML instances

inference data integration

Dataframe container w/ dictionary

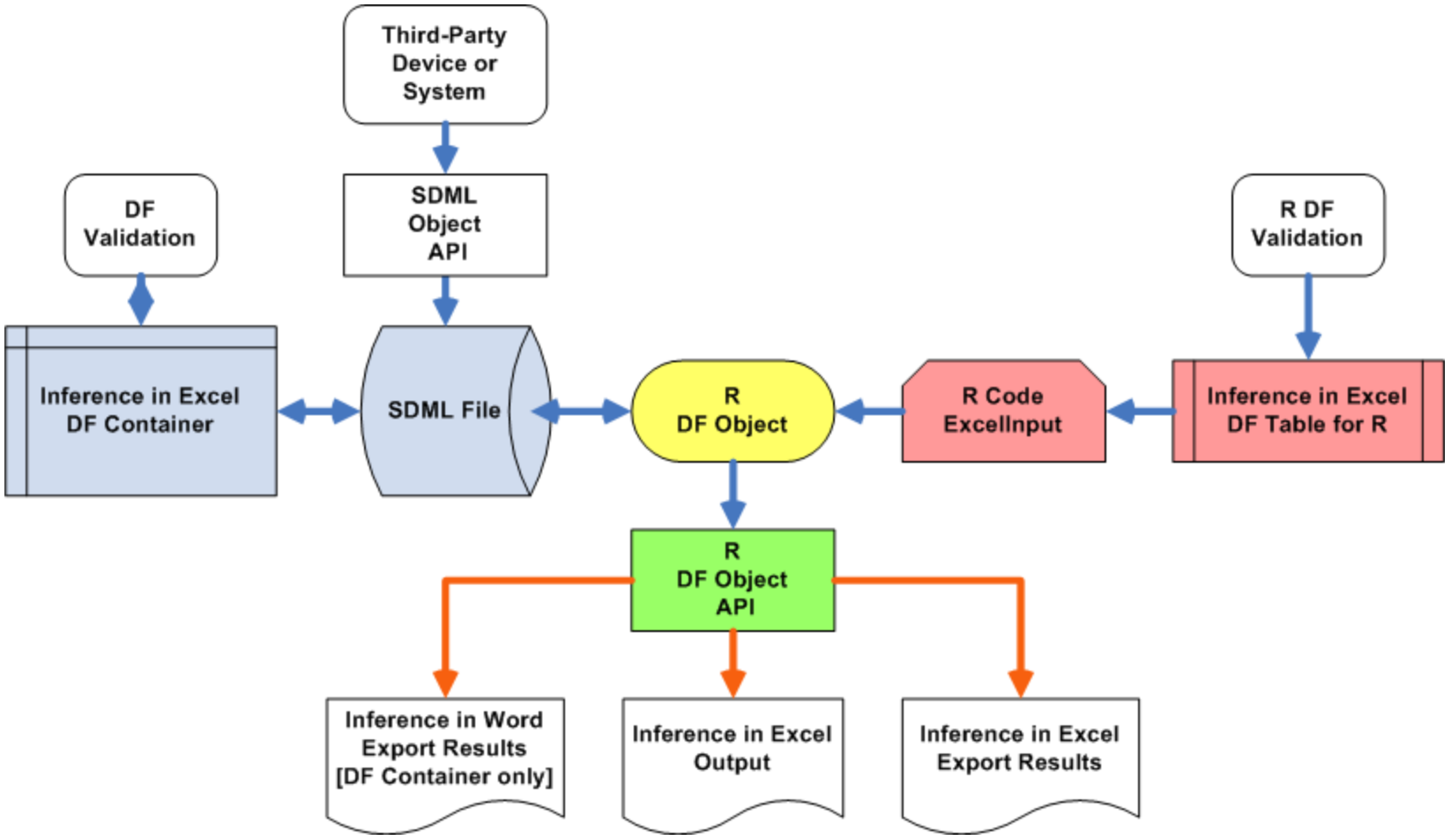
inference
from Blue Reference



	A	B	C	D	E	F	G	H	I	J
1	SDMLDataType		numeric-integer	categorical	numeric-integer	numeric-real	numeric-real	numeric-real	numeric-real	numeric-real
2	label		Run ID	day of block	Replicate ID	1-methyl-3-pyrr	Triethylamine T	Hold Time	Hold Temperat	Yield
3	units		NA	NA	NA	L/kg-LR	eq	hrs	degC	%
4	role		NA	factor	NA	factor	factor	factor	factor	response
5	column		1		2	3	4	5	6	7
6										
7	variable	index	runID	block	repID	pyrrolidone	teaHF	holdTime	holdTemp	yield
8		1	31	B4	23	5	1.335	25	10	75.82
9		2	3	B1	1	4	1.17	22	15	82.93
10		3	4	B1	3	4	1.17	28	15	88.07
11		4	22	B1	5	4	1.5	22	15	88.68
12		5	25	B1	7	4	1.5	28	15	93
13		6	28	B1	9	6	1.17	22	15	77.21
14		7	30	B1	11	6	1.17	28	15	83.6
15		8	5	B1	13	6	1.5	22	15	84.86
16		9	29	B1	15	6	1.5	28	15	88.71
17		10	16	B2	17	3	1.335	25	20	94.13
18		11	11	B2	18	7	1.335	25	20	89.94
19		12	14	B2	19	5	1.005	25	20	88.21
20		13	24	B2	20	5	1.665	25	20	93.11
21		14	1	B2	21	5	1.335	19	20	89.78
22		15	10	B2	22	5	1.335	31	20	94.61
23		16	7	B2	25	5	1.335	25	20	93.32
24		17	23	B2	25	5	1.335	25	20	92.32
25		18	19	B2	25	5	1.335	25	20	93.68
26		19	9	B2	25	5	1.335	25	20	93.27
27		20	27	B2	25	5	1.335	25	20	92.87
28		21	6	B2	25	5	1.335	25	20	92.96
29		22	15	B2	25	5	1.335	25	20	93.07
30		23	18	B3	2	4	1.17	22	25	94.04
31		24	17	B3	4	4	1.17	28	25	93.97
32		25	21	B3	6	4	1.5	22	25	94.3
33		26	12	B3	8	4	1.5	28	25	93.42
34		27	26	B3	10	6	1.17	22	25	92.99
35		28	20	B3	12	6	1.17	28	25	94.38
36		29	13	B3	14	6	1.5	22	25	94.26
37		30	2	B3	16	6	1.5	28	25	94.66
38		31	8	B5	24	5	1.335	25	30	93.25
39										

Dataframe objects: R .NET SDML

inference
from Blue Reference

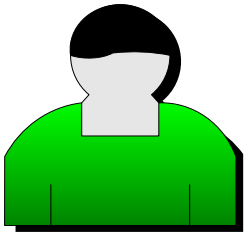


Test Driven Development

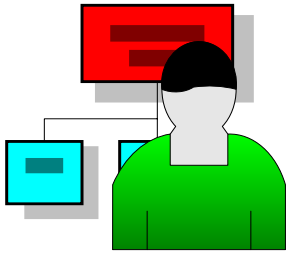
inference applications

Test-Driven Development: Workflow

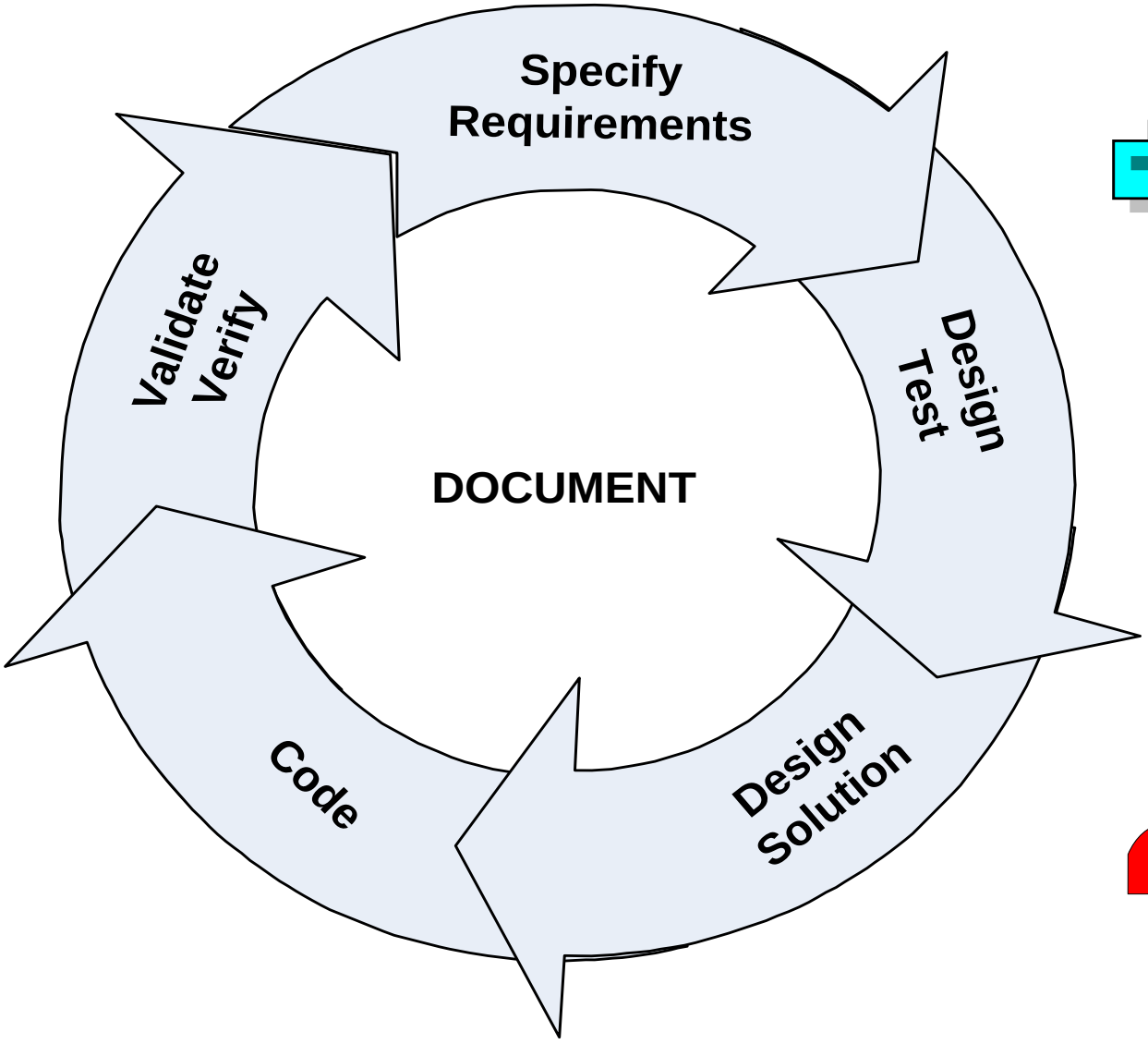
inference
from Blue Reference

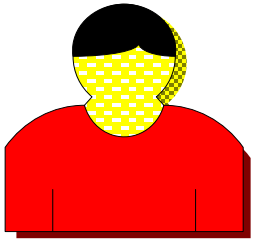


Developer/Analyst

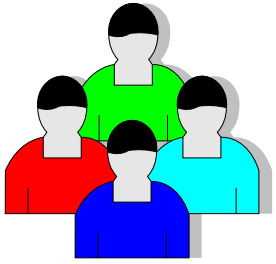


Business Systems





Domain Expert



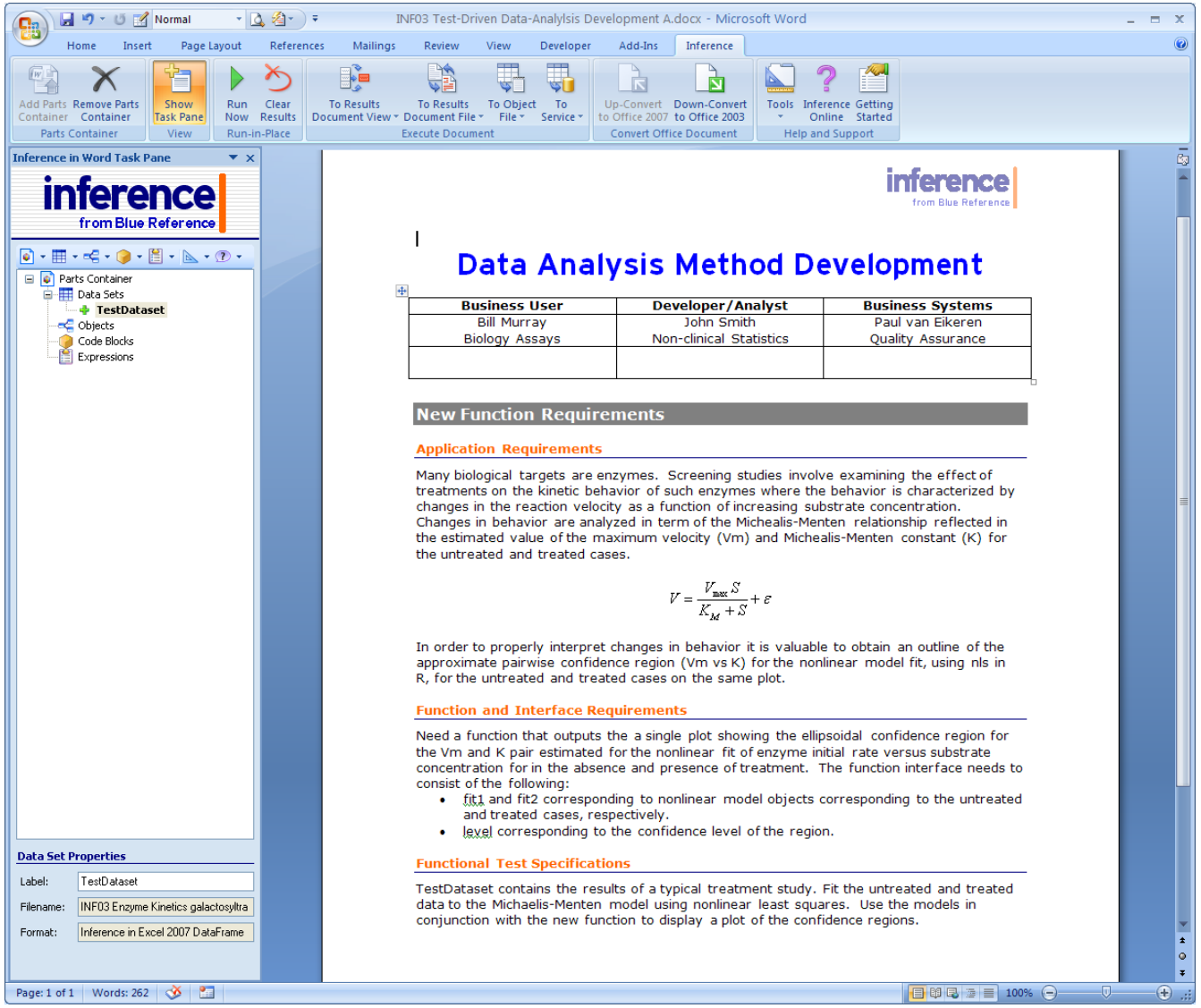
Customers

Test-Driven Development: benefits

- Benefits from integration
- Involves directed workflow and deliverables
  - Collect requirements and test data
  - Define test scripts
  - Specify code and run test scripts
  - Collect and report test verdicts (pass or fail)
  - Document entire process
  - Support collaboration among stakeholders
- Deliverables are basis for validation and verification

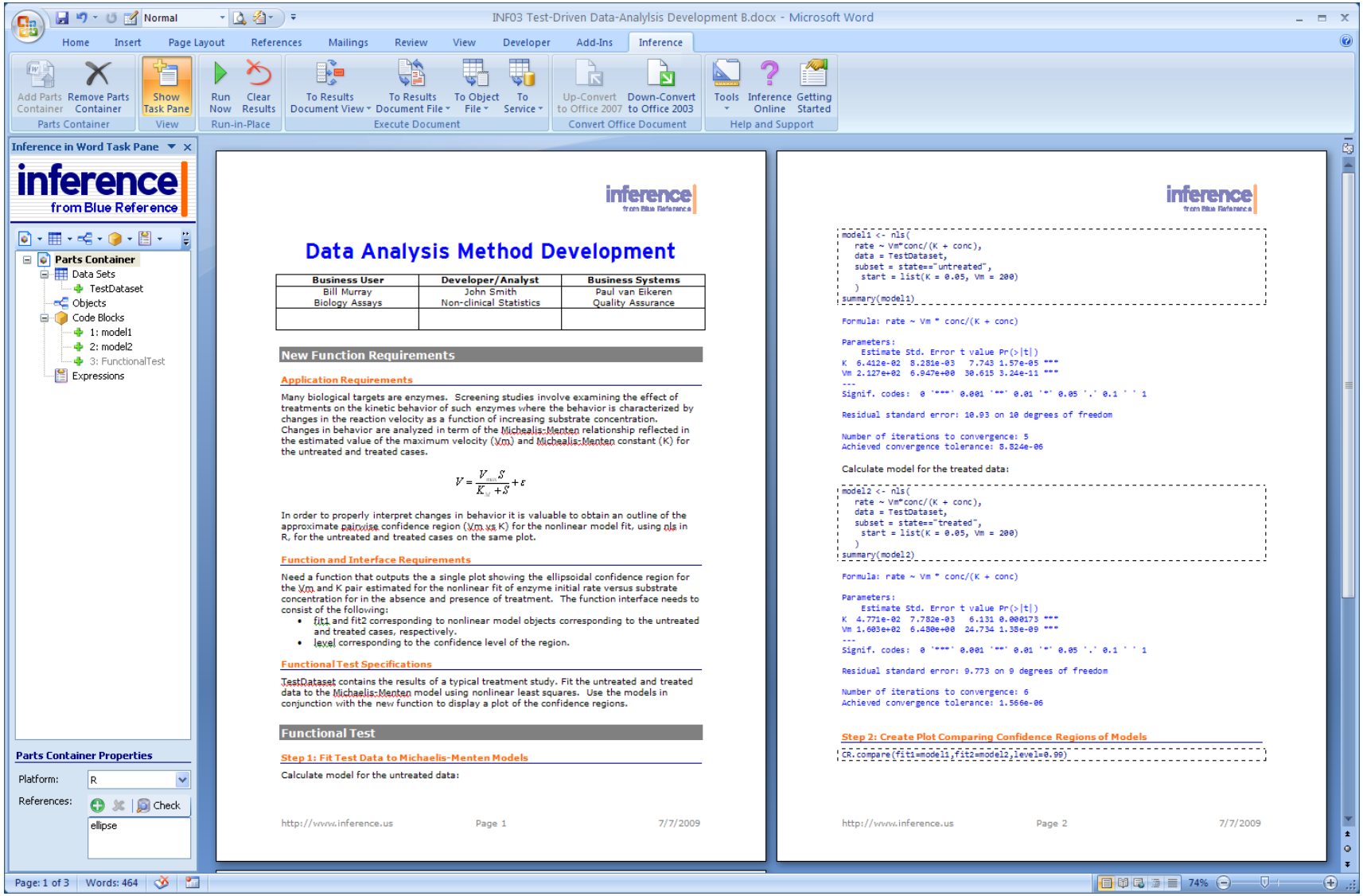
1. Requirements and Test Data

inference
from Blue Reference



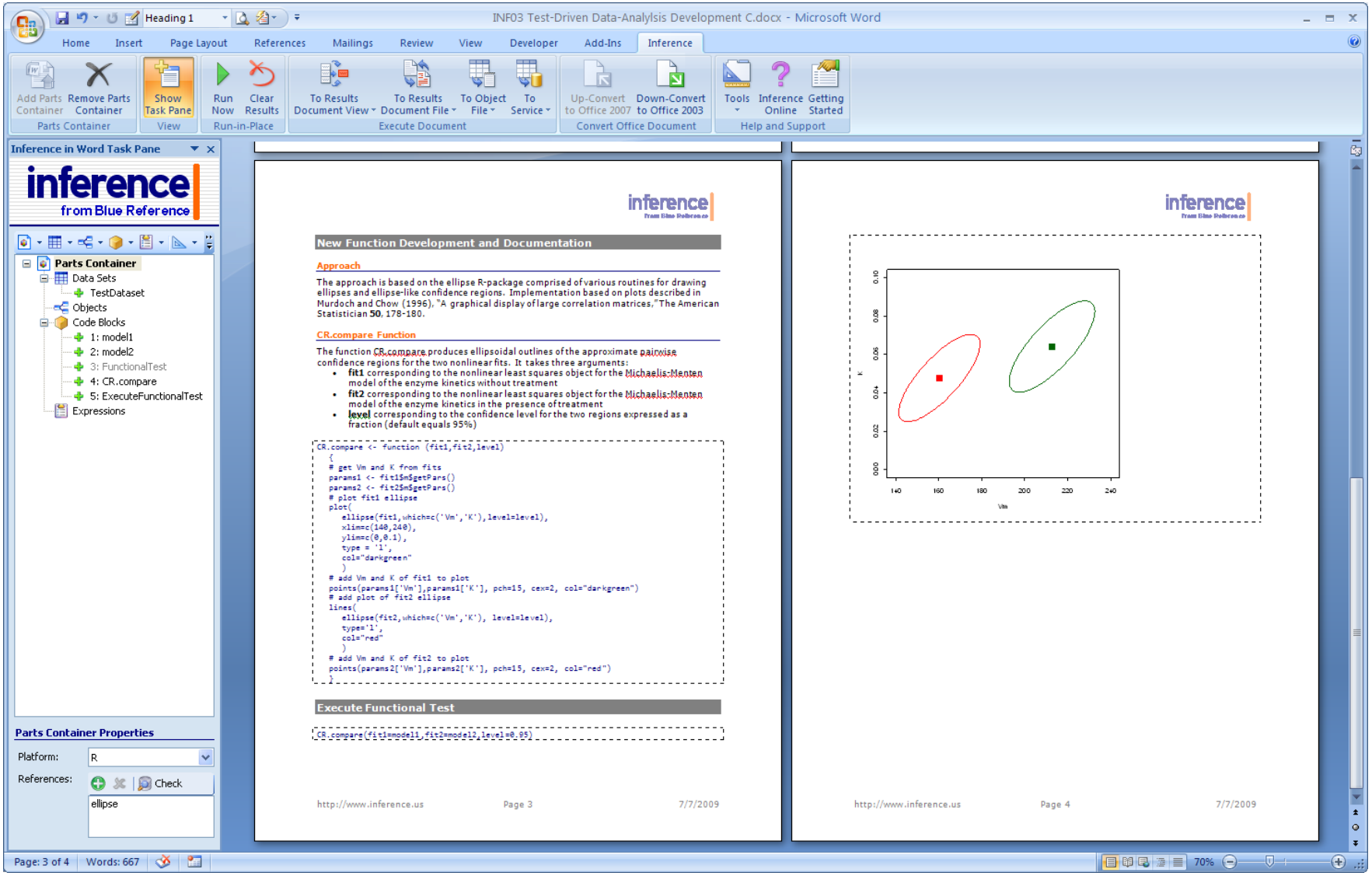
2. Define/Run Functional Tests

inference
from Blue Reference



3. Code, Test & Report Results

inference
from Blue Reference



Study Management

inference application

Preclinical Study Management

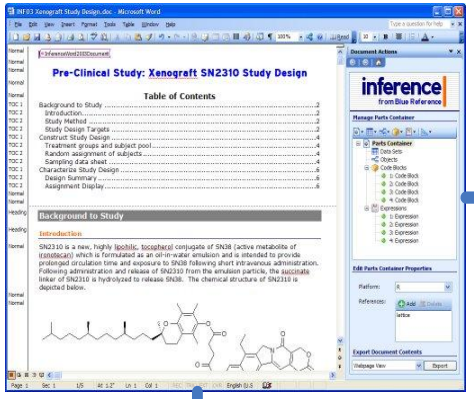
inference
from Blue Reference

- Test safety and efficacy of new drug candidates in animals before testing in humans
- Scientists rely on paper and Excel
- Paper and Excel inefficiencies:
  - lack of easy-to-use tools for study design
  - lack of collaboration support
  - need to transcribe and merge data
  - data not accessible in real time
  - manual creation of reports

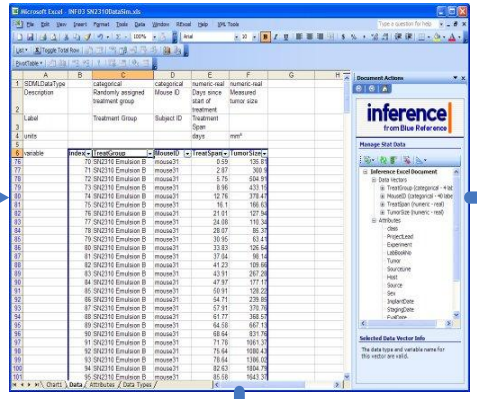
inference in Study Management

inference
from Blue Reference

Study Design

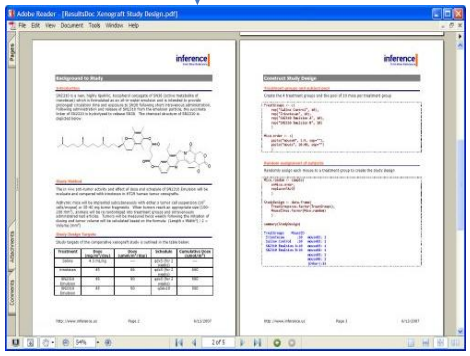


Design Data

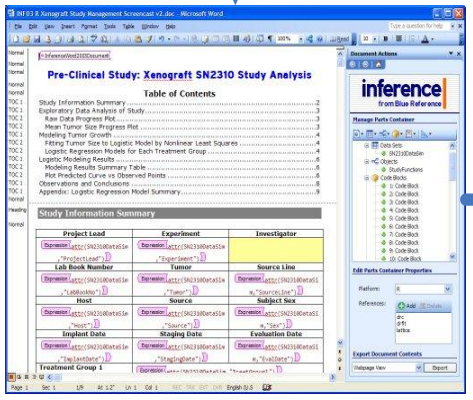


Data Export

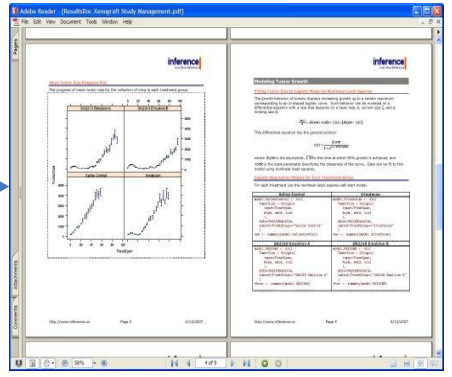
StatDataML
XML
Spreadsheet
Text File
Data Base



Design Report



Design Analysis

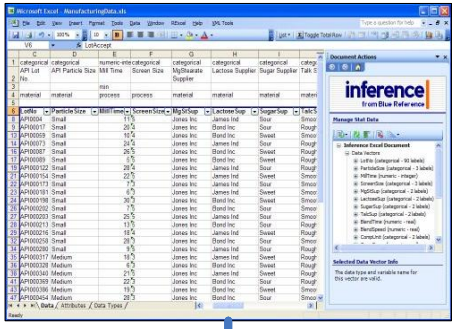


Analysis Report

inference for QbD study management

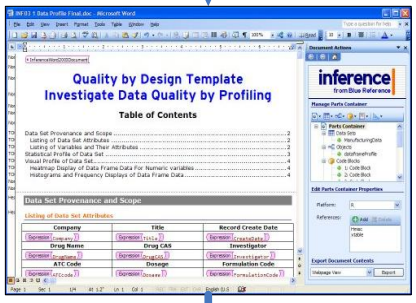
inference
from Blue Reference

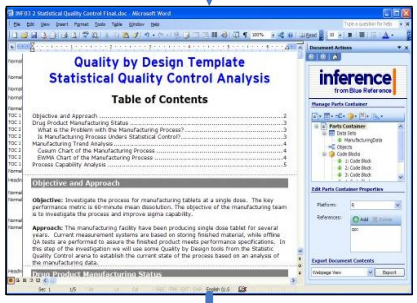
Inference Excel

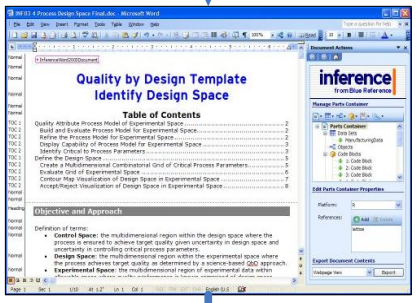


Manufacturing Data 1

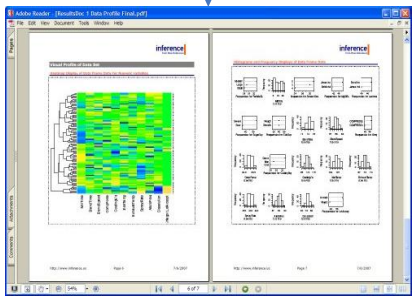
2 3 Inference Word Templates 4 5



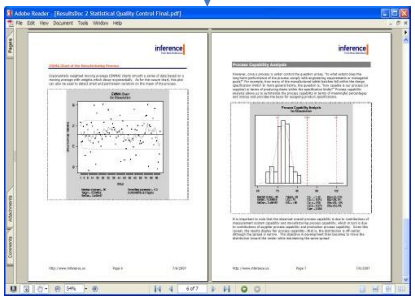




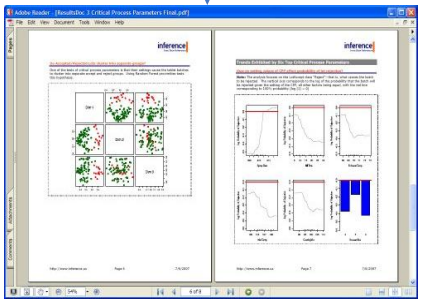
Inference Results Documents



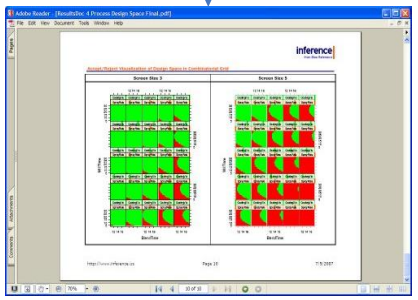
Data Profile



Statistical Quality Control



Critical Process Parameters



Design Space

take-home message

inference for R value

inference
from Blue Reference

- Obtain increased returns on your investments in R training
- Extend R data-analysis capability to non-expert users
- Manage R data-analysis development, execution, collaboration and publication as a document-based business process
- Provide R data-analysis service as software
- Deploy “one-button” R data-analysis solutions
- Reuse R data-analysis solutions

For additional information contact:
Paul van Eikeren, Ph.D.
Paul.van.Eikeren@BlueReference.com

www.InferenceforR.com