

Lifecycle Processes to Insure Quality of Blaise Interview Data

Linda Gowen, Westat

Pat Clark, Westat



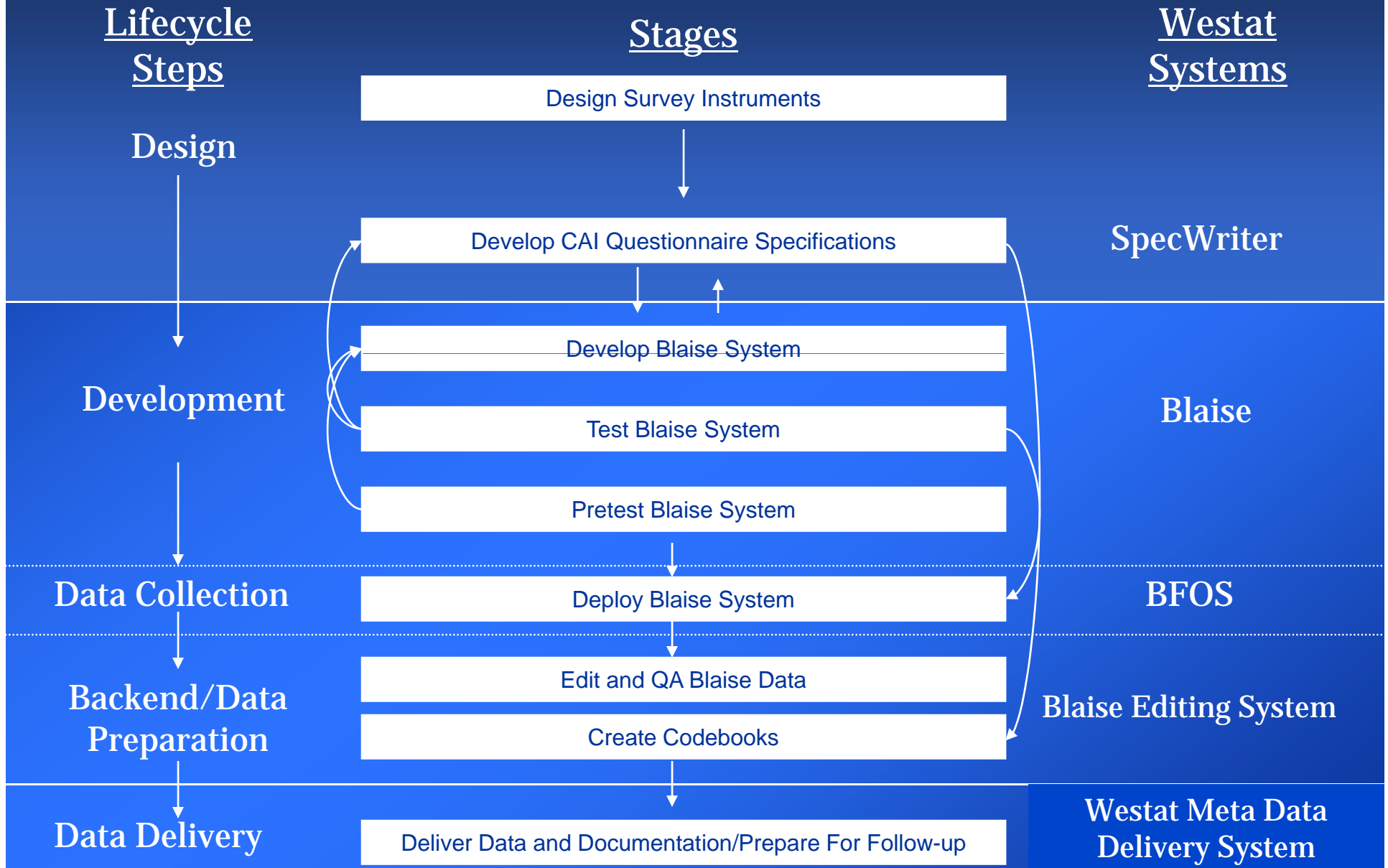
Lifecycle Processes to Insure Quality of Blaise Interview Data

- This presentation reviews some of the Westat tools used to facilitate processing at the different stages throughout the lifecycle of a Blaise project, with a focus on the impact on data quality.

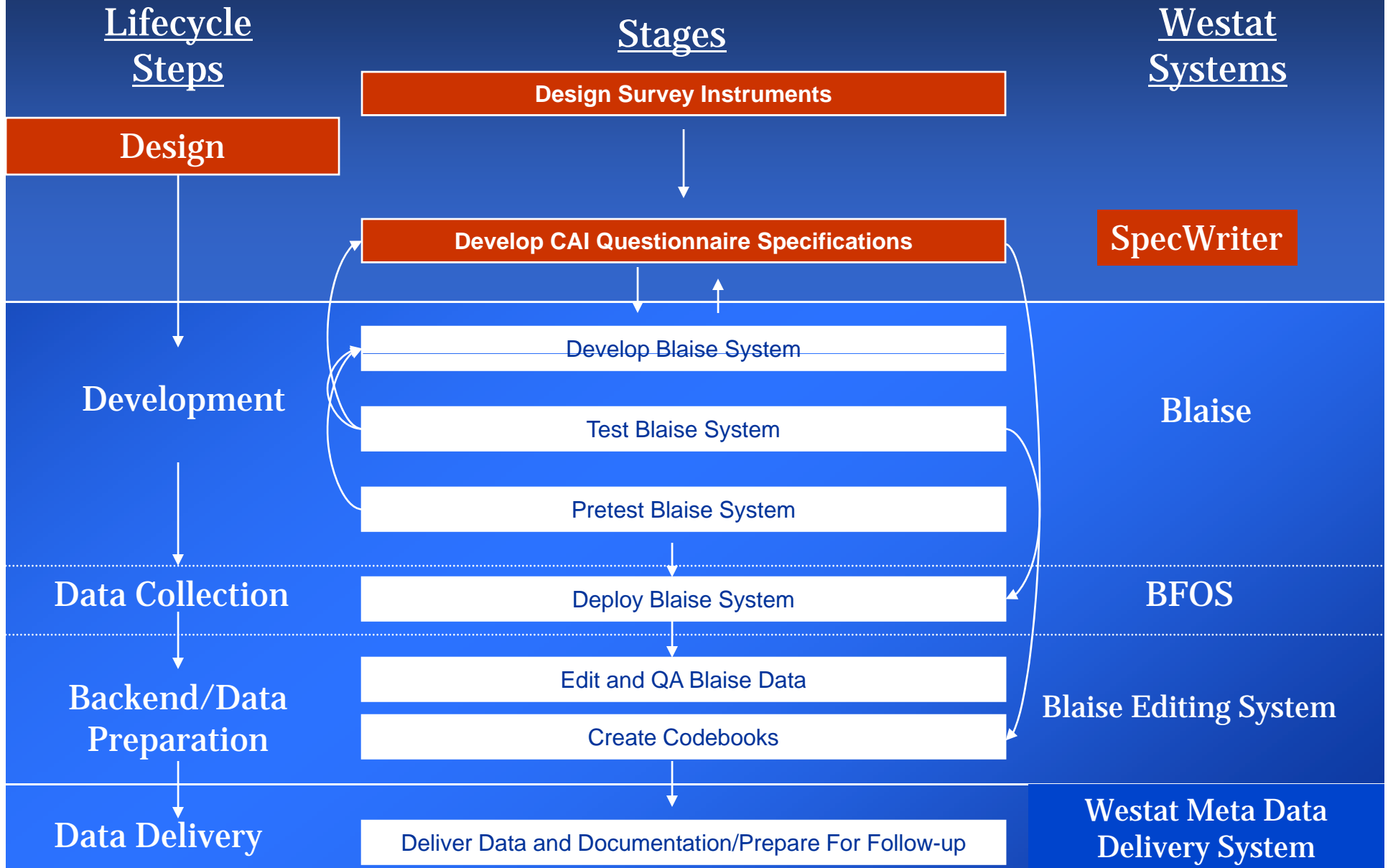
Survey Lifecycle Overview

- The major lifecycle steps are:
 - Design
 - Development
 - Data Collection
 - Data Preparation
 - Data Delivery
- Westat tools are used at every stage to improve data quality.

Lifecycle Processes



Lifecycle Processes



Stage 1 – Requirements Analysis

- The goal of this stage is to define both the Blaise and non-Blaise requirements for the system.
- It is helpful to prepare a project data flow diagram showing:
 - All sources providing data into the Blaise Systems
 - All data required to feed external systems
 - All data transformations

Stage 2 – CAI Specifications

The CAI specifications document includes the “complete” hard copy questionnaire typically needed for regulatory compliance, with additional detailed instructions, necessary for programming the CAI system.

Question# Item#	Field Name Question Text
Q23.1 RSY1005	Cough_AM Now I am going to ask you about respiratory symptoms that you may have experienced in the past year. Do you usually cough on waking up, or first thing in the morning? 0. NO 1. YES Don't Know. Refused.
Q23.2 RSY1010	Cough_PM Do you usually cough during the rest of the day or at night? 0. NO 1. YES Don't Know. Refused.
Q23.3_Box RSY1012	IF Question Q23.1(Cough_AM) <> Yes AND Question Q23.2(Cough_PM) <> Yes GOTO Question Q23.5(CoughPhlegm_AM)
Q23.3 RSY1015	PastYrCough_Plus3Mos During the past 12 months, have you had this cough on most days for three months or more? 0. NO 1. YES Don't Know. Refused. Refused. Refused.

SpecWriter Output

Questionnaire

Section: DIA Section 14 - Diabetes

Question# Item#	Field Name Question Text
Diab_DSP D/A1001 Size: 255	DiabDisp Diabetes Display - Gender Clause [_____]
ASSIGN_DiabDsp D/A1003	<div style="border: 1px solid black; padding: 5px;"><p>IF Question Gender(Gender) = Male ASSIGN Question Diab_DSP(DiabDisp) := " " ELSE ASSIGN Question Diab_DSP(DiabDisp) := ", other than when pregnant"</p></div>
Q14.1 D/A1015	Diabetes Have you ever been diagnosed with @Bdiabetes or high blood sugar@B ^DiabDisp? 0. NO(Skip to Question Q15.1) 1. YES Don't Know. (Skip to Question Q15.1) Refused. (Skip to Question Q15.1)
Q14.2 D/A1010 Size: 3	Diabetes_Age How old were you when you were first diagnosed with diabetes or high blood sugar? [][][][] [Range:0 to 120] Don't Know. Refused.
Q14.2 FNIT D/A1012	Hard Edit: [Question Q14.2 Diabetes_Age] <= Question Resp NTAge(Respondents_ nt_Age)]; If not true display message. "Age when diagnosed with diabetes or high blood sugar must be less than or equal to current age."
Q14.3 D/A1015	Diabetes_Med Do you currently take any prescribed medicines for this condition? 0. NO(Skip to Question Q15.1) 1. YES Don't Know. (Skip to Question Q15.1) Refused. (Skip to Question Q15.1)
Q14.4 D/A1020	Insulin Do you currently take insulin? 0. NO 1. YES Don't Know. Refused.

Display Field Defined

Display Assignment Logic Box

Skip Pattern from Answer Response

Question Number

Item #

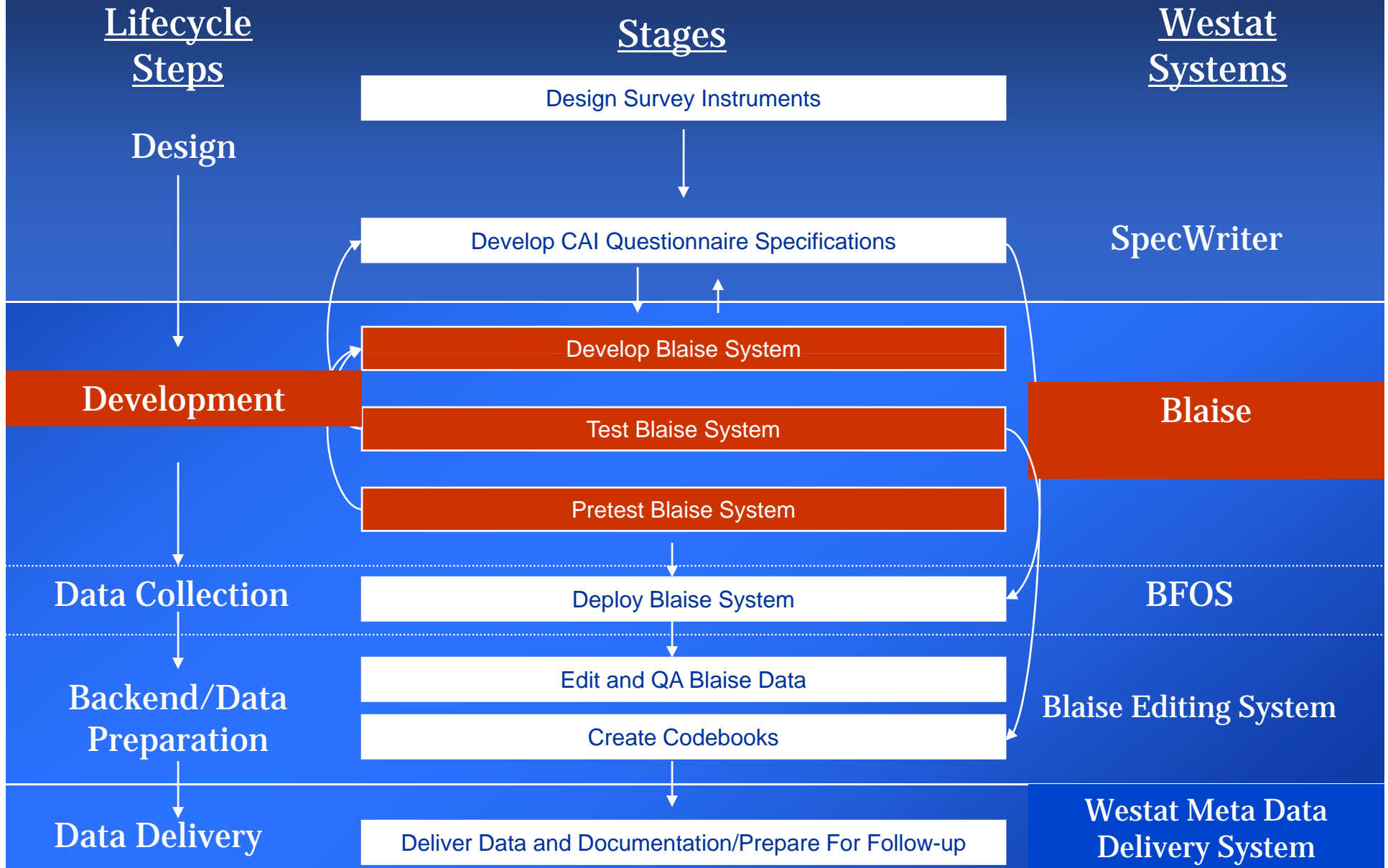
Display Field Reference

Numeric Range Declaration

Edit



Lifecycle Processes



Stage 3 - Blaise Development

- Programmers reference documents
 - Project data flow diagram
 - CAI specifications

Blaise Development Key Considerations

- Proper use of data to be used as foreign keys in the creation of blocks/tables
- Review of complicated programming
- Testing and QC at all stages of instrument development including verification of the data integrity after both forward and backward navigation
- Declaring fields to be compatible with external systems
- Review relationship of question numbers to Blaise variable names
- Save needed intermediate variables

Blaise Development Operational Considerations

- Identify data for external systems performing quality assurance and monitoring.
- Advance planning can assure the database structure provides for the easy extraction of data.

Stage 4 - Internally Testing the Blaise System

- A specialized testing team ensures a level of quality
- CAI Specifications are tested
- Issue Tracking
- Feature by feature validation
- Functional testing
- Testing is done using all systems
- All testing is based on workflow
- Regression testing

Stage 5 – Pretest for Field Readiness

- Testing is also performed by the project staff, data management staff, and field staff.
- Role plays should simulate data collection scenarios.
- The goal is to have each real condition tested for all the responses available in the instrument.
- Files with the delivery layout are built from the mock data and frequencies run to verify that the system is capturing the data to meet the research and operational objectives.

100% 8.5 B I U [List of icons]

Courier New 8.5 B I U [List of icons] CATS Normal - [List of icons]

[HAVE TRAINEE READ]:

WaterSource. Let's talk about your current home in Shaftsbury. You said that you lived here from 1975 to the present. During this period, what source or sources supplied water to this home?

[INTERVIEWER: SHOW CARD RS-1. SELECT ALL THAT APPLY.]

[INTERVIEWER: RECORD WATER SOURCE (PUBLIC OR PRIVATE) ON CALENDAR.]

- A PRIVATE WELL OR SPRING..... (10)
- A COMMUNITY WELL OR SPRING SERVING AT LEAST 25 PEOPLE OR 15 HOUSEHOLDS..... 11
- A PUBLIC OR MUNICIPAL WATER SUPPLY..... (12)
- ANOTHER SOURCE OF WATER (SPECIFY)..... 91

[RESPONDENT]: "I've got city water and I also have a private well."

NumberWellsSprings. How many different private wells or springs supplied water to this home?

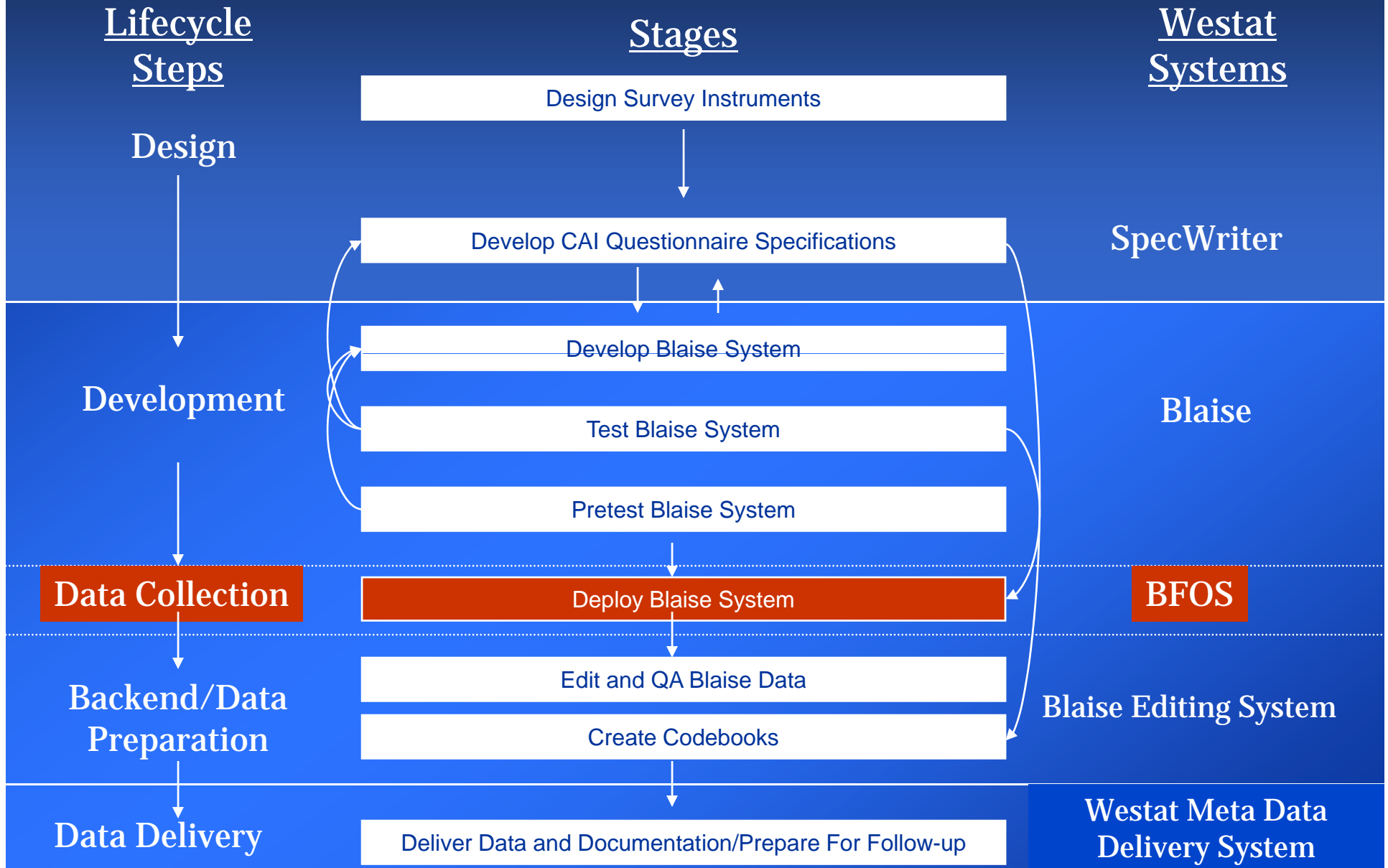
[INTERVIEWER: ENTER "5" FOR FIVE OR MORE PRIVATE WELLS OR SPRINGS.]

| 1 |

[RESPONDENT]: "One"



Lifecycle Processes





Cases:

Case ID	Respondent Name	Disposition	Disposition Date
901304B	George Washington	EL-Elgibile	Thu. 5/25/2006 10:16 AM
901309B	John Doe	EL-Elgibile	Thu. 5/25/2006 12:38 PM
901310B	Mary Clark	EL-Elgibile	Wed. 5/31/2006 2:32 PM
901403A	Dennis Miller	EL-Elgibile	Wed. 5/31/2006 3:13 PM
901509A	David Williams	EL-Elgibile	Fri. 6/2/2006 11:07 AM

Tasks:



Screener	Baseline	Quarterly 1	Quarterly 2	Quarterly 3	Quarterly 4	Quarterly 5	Quarterly 6	Quarterly 7	Followup
CO-Complete									

BFOS

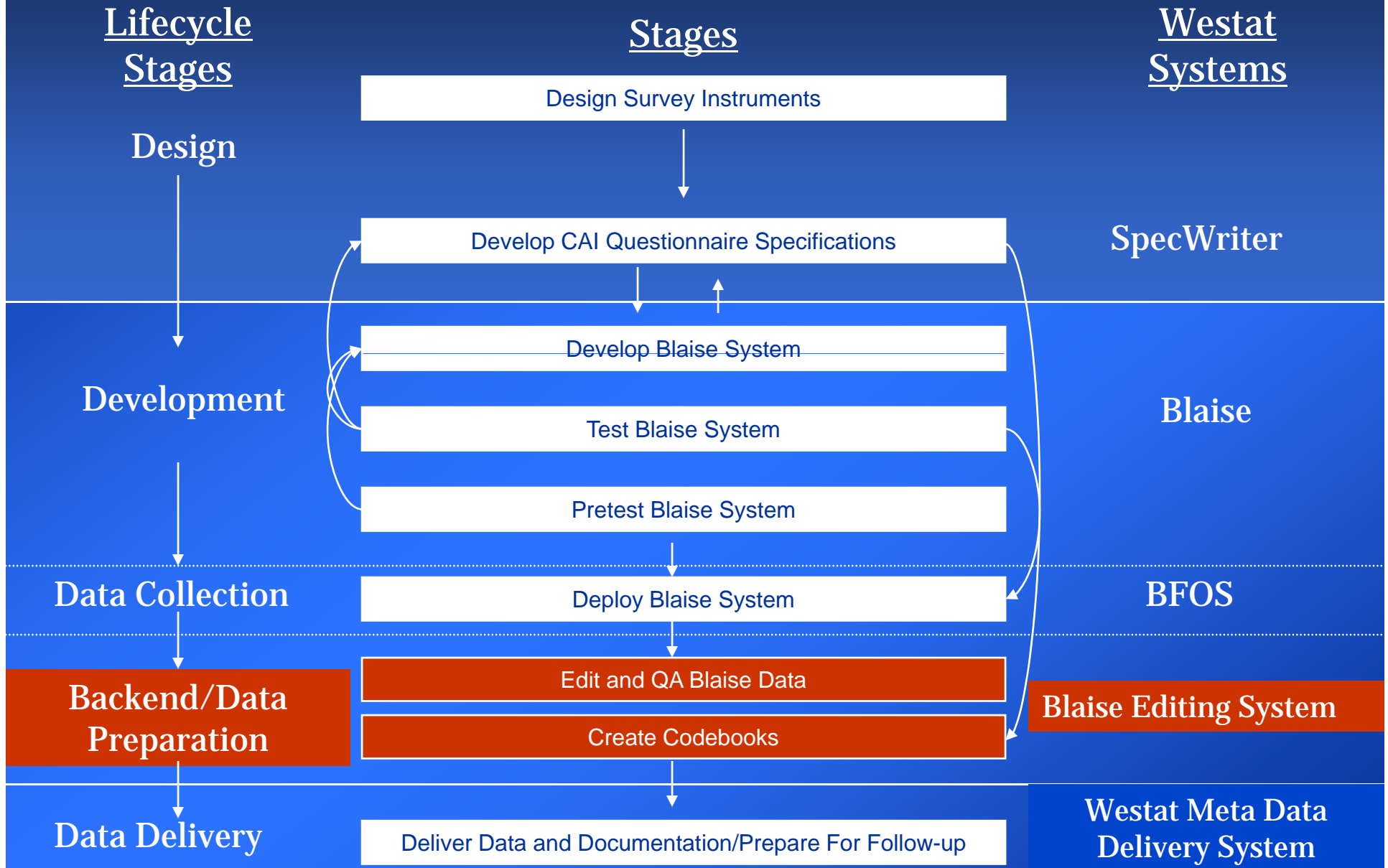
BFOS provides projects with:

- Interviewer Management System - A laptop-based application that allows interviewers to review and change assigned case status, launches the Blaise CAPI instrument, and synchronizes the management and CAPI data with the home office.
- Supervisor Management System - This is the web-based application used by field supervisors to view and manage the work of interviewers, including case assignment and transfer between interviewers, review and adjustment of case status, and other administrative and reporting functions.
- Home Office Management System - This provides functions needed by home office staff to load and manage the study sample information and to maintain information about the field staff and their assignments.

Planning Prevents Unexpected Results



Lifecycle Processes



Stage 7 – Editing Blaise Data

An array of systems are available to support Blaise editing.

- Editing system
- Extract system
- Comment system
- Data decision log
- Data and interview validation
- QA reconciliation systems
- Study management systems
- Metadata delivery system

Stage 7 – Editing Blaise Data

Basic steps of the editing process are:

- Combine data into a master database
- Maintain Blaise interview in interviewing and data editing modes
- Display comments
- Editors may make changes to data.
- Decisions documented.
- Other, specify fields coded.
- Frequencies reviewed
- Reconciliation completed.

Editing : Database (Access 2000 file format)

Open Design New [Icons]

Objects

- Tables
- Queries**
- Forms
- Reports
- Pages
- Macros
- Modules
- Groups
- Favorites

Name

- CompleteIDs Without Mat...
- CompleteIDs_Without_M...
- CompletesToAdd
- CreateProstPArtd_aps
- Find duplicates for PIDs
- Fix_PIDSNotAddedToEd
- Ineligible_Completes
- PhaseII_Prostate_Canc
- PIDs Without Matching f
- PIDs_All Without Matchi
- PIDsInEditingSystem
- Update_CompleteIDs

Blaise 4.7 Data Entry - \\rk25\vol2501\ahs\Phase3\Blaise\EditingSystem\EditData\EditData.d...

Forms Answer Navigate Options Help

ID: 100251P, NAME: , VERSION:

During the past 12 months, have you had this cough on most days for three months or more?

0. NO
 1. YES

Q23.1 Cough_	1	Yes
Q23.2 Cough_	0	No
Q23.3 PastYr	0	No
Q23.4 Cough_	15	
Q23.5 Cough	0	No
Q23.6 Cough	1	Yes
Q23.7 PastYr	0	No

Old 67/79 Modified by rules Clean Navigate AHSPhase3

Remark

more prevalent in spring and fall

Save Cancel Help

Data Decision Log

Data Decision Log *
Tuesday, June 26, 2007

Study Name: Agricultural Health Study
Delivery: Filtered
firstpage

Decision #	Person ID	Time In Study	Source/Form ID	Item ID	Date/Element Name	Description of Issue	Inquiry Source	Original Value	New Value	Decision	Reason	Date of Decision	Approved By	Link to Type Code From Study Decision	Item Path Name	Item Deliverable
948	100008P		PST - V1	Q4.7	Lifetime Years Per Year	SEVERE YEARS 4, 1 DAY A YEAR ROUND UP LULU TRA YEARS 8 AND 7 DAYS A YEAR BKECP 2 YEARS, 15, 7 DAYS	DM Review	12/08/2005	postcode	Review - 2	Change	10/1/2006	Pat O'Leary			<input type="checkbox"/>
947			RSP - V1	Q13.7	Age/Risk of death Farmer's Lung	DR SAID HE MIGHT HAVE HAD IT	DM Review	12/08/2005	non-postcode	Review - 40	Change	10/10/2006	Coding Rules			<input type="checkbox"/>
946			ALR - V1	Q17.1	Ever medically diagnosed hay fever, allergic rhinitis	HAS A REACTION TO GRAIN DUST, BUT HAS NEVER BEEN CHECKED	DM Review	12/08/2005	No Action Necessary	NO	Leave As Is	0/13/2006				<input type="checkbox"/>
945			RBY - V1	Q23.4	Total years R had this cough	ERROR	DM Review	12/08/2005	No Action Necessary	REFUSED	Leave As Is	0/8/2/2006				<input checked="" type="checkbox"/>
940			PRT - V1	Q27.2	PSA level normal	HAVENT GOTTEN RESULTS YET	DM Review	12/08/2005	No Action Necessary	DONT KNOW	Leave As Is	0/7/7/2006				<input type="checkbox"/>
1070.0	100008P		PST - V2	Q4.7a	Postcode, Variable m1	DUAL & BROADSTRIPE MIGHT HAVE BEEN TOGETHER	DM Review	11/08/2005	postcode	Break	Change	12/13/2006	Joe Cade			<input type="checkbox"/>
1008.0	100008S		NSY - V2	Q24.4	Handwriting smaller than 4 mm on size	SHE SAID "NOT MUCH" BUT HER RIGHT HAND IS IN A BRACE RIGHT NOW DUE TO SURGERY ON CARPUL TUNNEL	DM Review	05/17/2007	No Action Necessary	DONT KNOW	Leave As Is	0/30/4/2007				<input type="checkbox"/>
1023.7	100020P		I20 - V2	Q4.6	Lifetime Years	Incorrect Postcode(Q4.6)	No/Used	11/08/2005	postcode	INCORRECT PESTICID E	Change	12/1/2006	Coding Rules			<input type="checkbox"/>
1023.8			I20 - V2	Q4.6	Lifetime Years	Incorrect Postcode(Q4.6)	No/Used	11/08/2005	postcode	INCORRECT PESTICID E	Change	12/1/2006	Coding Rules			<input type="checkbox"/>

Tuesday, June 26, 2007

* Due to space limitations, two columns are not shown: "Comment" and "Date of Last Update"

Page 1 of 122

Frequency Review

- As a final check of the data files themselves, the data are extracted from the final edited tables and frequency reviews are performed.

Quality Assurance

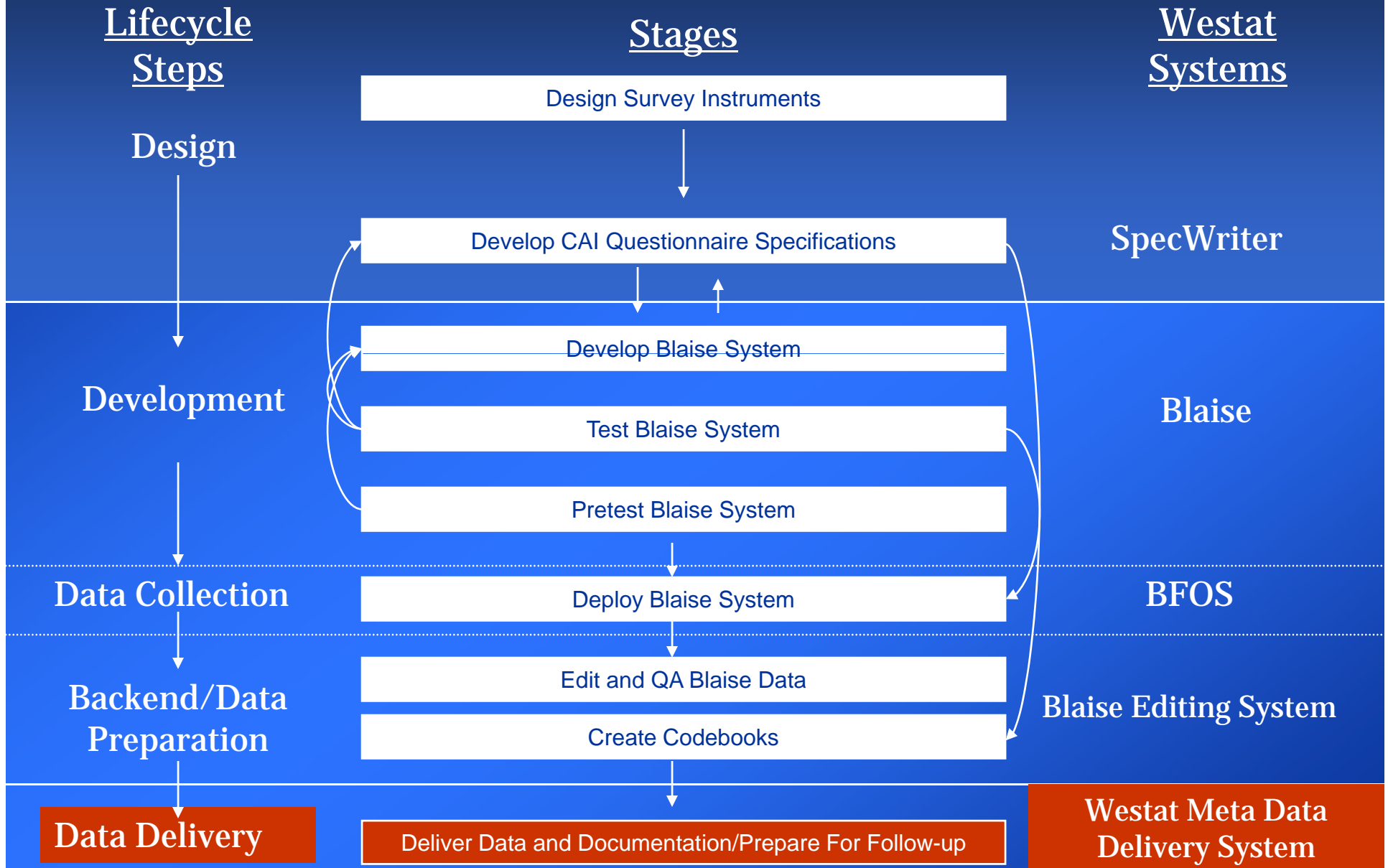
- Redundant data.
- QA system
 - Report discrepancies
 - Secure
 - Reports on Editing system data, SMS data, other project source data
- Documentation of resolutions recorded in Data Decision Log.

Stage 8 – Create Codebooks

- The codebook is an important part of documentation.
- Our codebook generation system was designed around the metadata output from the Blaise instrument.
- Data elements are versioned.
- All derived data elements have the associated derivation details present in notes.

Field Name	Data Element Name and (DE Number)	Element Source	Source Details			Data Type/Length	Position or Column Start: End	Format
			Form ID	Form Section	Item ID			
CAPI_AGE (7672)	CAPI	CAPI				Number/ 3		
Item Text: CAPI Age: Calculated from the DOB recorded and verified at the time of the interview								
<u>Value</u>	<u>Definition (*Skip)</u>						<u>Frequency</u>	
0-29	= Less than 30						10	
30-100	= Over 30						387	
							Total = 397	
Derivation Details: CAPI_Age = Consents.StudyCAPIDate - DOB(Demographics.DOB_month,,DOB_Day, DOB_year)								
BOX 1 IF PARTICIPANT IS LESS THAN 30 YEARS AT TIME OF INTERVIEW, GO TO WF_CurrentWeight.(CAPI_AGE < 30) Go to WF3(WF_CurrentWeight)								
WF_HeightAge21Ft (2760)	CAPI	WF.v7	Weight Fluctuation	WF-2.v1		Number/ 3		
Item Text: How tall were you at age 21? [INTERVIEWER: ENTER FEET.]								
<u>Value</u>	<u>Definition (*Skip)</u>						<u>Frequency</u>	
4-6	= FEET						383	
9	= DONT KNOW (*Skip WF_HeightAge21Inch - WF_HeightAge21Inch (Code null))						4	
							Total = 387	
WF_HeightAge21Inch (2761)	CAPI	WF.v7	Weight Fluctuation	WF-2.v1		Number/ 3		
Item Text: HOW TALL WERE YOU AT AGE 21? [INTERVIEWER: ENTER INCHES.]								
<u>Value</u>	<u>Definition (*Skip)</u>						<u>Frequency</u>	
0	= NONE						45	
1-11	= INCHES						338	
							Total = 383	

Lifecycle Processes



Stage 9 – Data Delivery

Westat has developed a Data Delivery Metadata System which supports delivery associated with datasets, such as:

- Raw data
- Research datasets
- Analytic datasets
- Restricted use datasets, and
- Public Use files.

Data Delivery Metadata System



Study:

Delivery Type:

Delivery Name:

Delivery Version:

Delivery Dataset:

Category:

Search For:

In: Data Element
 Label

Go

Results Page: 981 records found.

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) ...

Study	Delivery Info	Dataset Info	Category	Data Element Label	Data Element Name	Source Detail
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_Antibiotics	BL_Antibiotics	Bio-Capi Quexv1 BC_419V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_BloodDrawPosition	BL_BloodDrawPosition	Bio-Capi Quexv1 BC_351V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_BloodDrawPositionRecode	BL_BloodDrawPositionRecode	Bio-Capi Quexv1 BC_353V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_BloodDrawPositonSpec	BL_BloodDrawPositonSpec	Bio-Capi Quexv1 BC_352V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_BloodThinners	BL_BloodThinners	Bio-Capi Quexv1 BC_309V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_CommentsBlood	BL_CommentsBlood	Bio-Capi Quexv1 BC_424V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_DateBloodDay	BL_DateBloodDay	Bio-Capi Quexv1 BC_349V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_DateBloodMonth	BL_DateBloodMonth	Bio-Capi Quexv1 BC_348V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_DateBloodYear	BL_DateBloodYear	Bio-Capi Quexv1 BC_350V1
NESEH	NESEH December 30 2005 delivery Version 1	CAPI Bio-Capi v1	CAPI Bio-Capi	BL_DrinkBeer	BL_DrinkBeer	Bio-Capi Quexv1 BC_358V1

[1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) ...

Summary

- Managing aspects of the processing extending across multiple lifecycle stages is critical to data quality.
- Feedback from all aspects of the survey need to be considered, and iterative processing loops are necessary.
- Lifecycle tools can be implemented in multiple project environments and customized to the needs of individual surveys.
- Active participation from all team members assure that differing perspectives are represented.
- Quality control becomes continuous from planning through delivery, ensuring accurate and efficient data collection and processing.