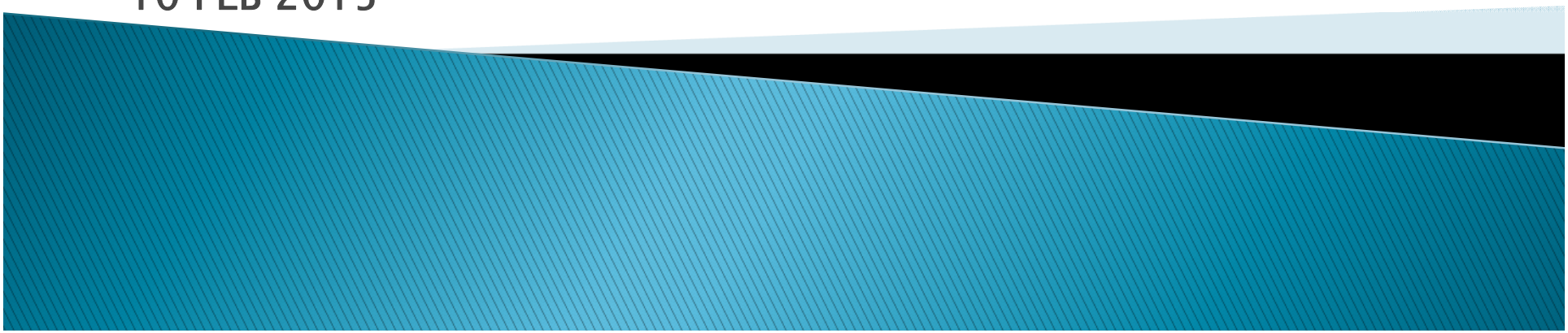
Interpreting and Using the Validation Results from Automated Tools

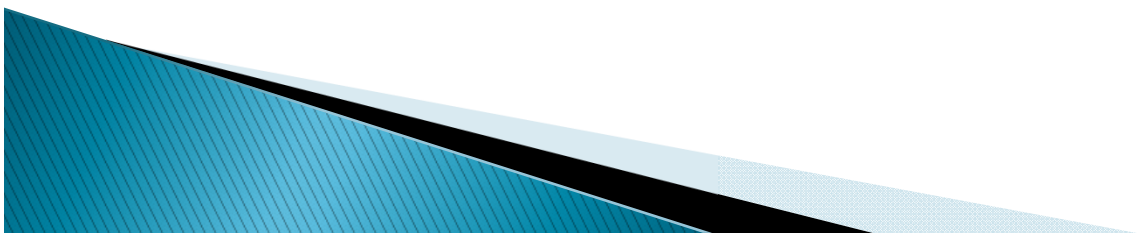
David Borbas
Sr. Director Data Management
Jazz Pharmaceuticals
Bay Area PharmaSUG
10 FEB 2015



Disclaimers

The opinions expressed in this presentation are not the official views or policies of Jazz Pharmaceuticals

- Where are we today?

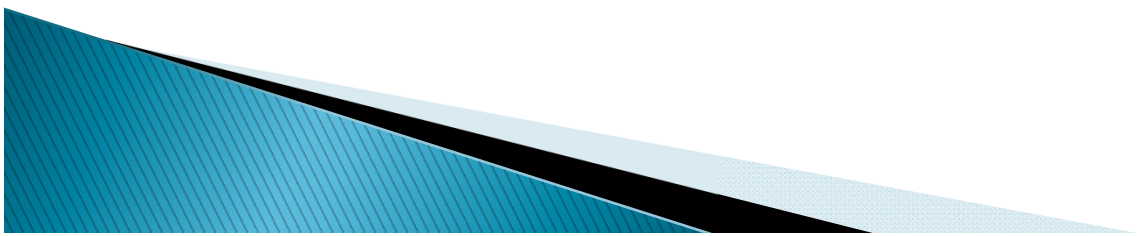


Current Environment

1) Submitting data electronically to the FDA and other regulatory agencies
2) eSubmission is in your future

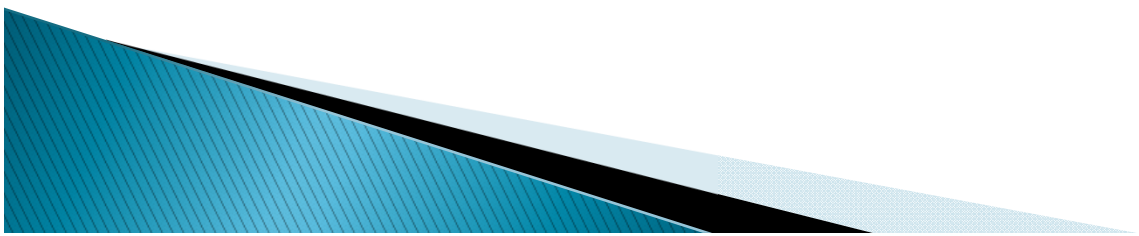
==========

- Sas Transport files, define.xml, blankcrf, reviewers guides and a study data standardization plan is required for eSubmissions of study data



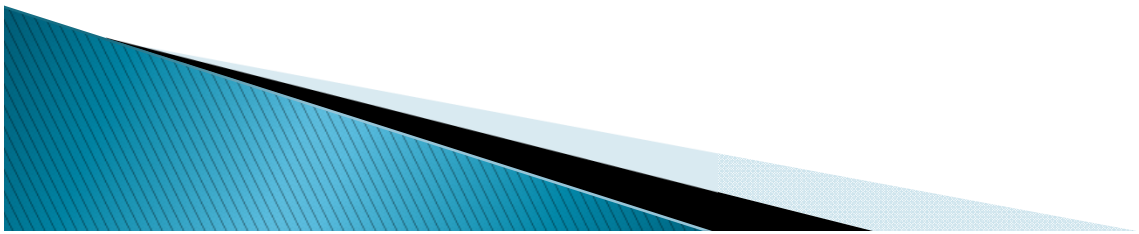
Submission Deliverables

- Not just datasets and define.pdf anymore
- Accuracy demands cross checks between
  - Datasets and define
  - Define and blankcrf
  - Data Standards plan, datasets and define
  - Between ADaM and SDTM



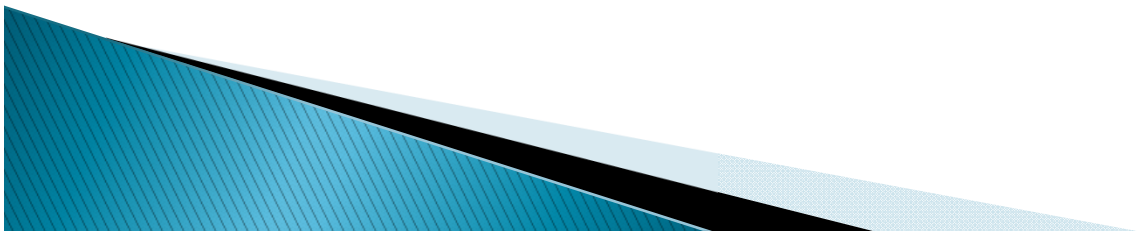
Required Submission Deliverables 2015

- Submission level
  - Data standardization plan
- Study level – SDTM and ADaM
  - SAS transport files
  - Define.xmls
  - Reviewers guides
- SDTM
  - blankcrf (annotated)



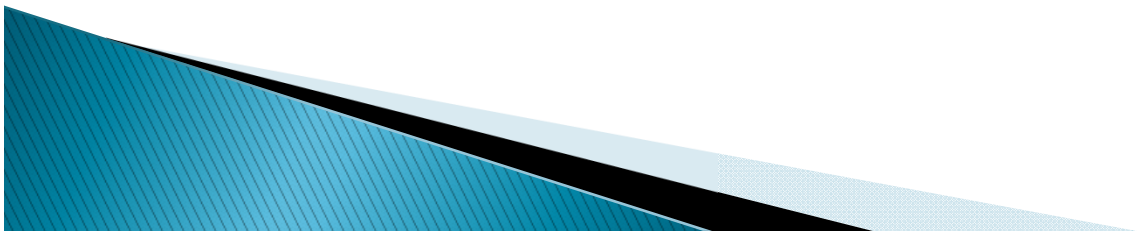
Submission Deliverables 2015 – 2

- There should be conformance to standards
- Integrity across and within each deliverable
- We want to pass the JumpStart Data Fitness test that FDA Computational Sciences will apply to NDA and BLA filings



What is Define.xml? -1

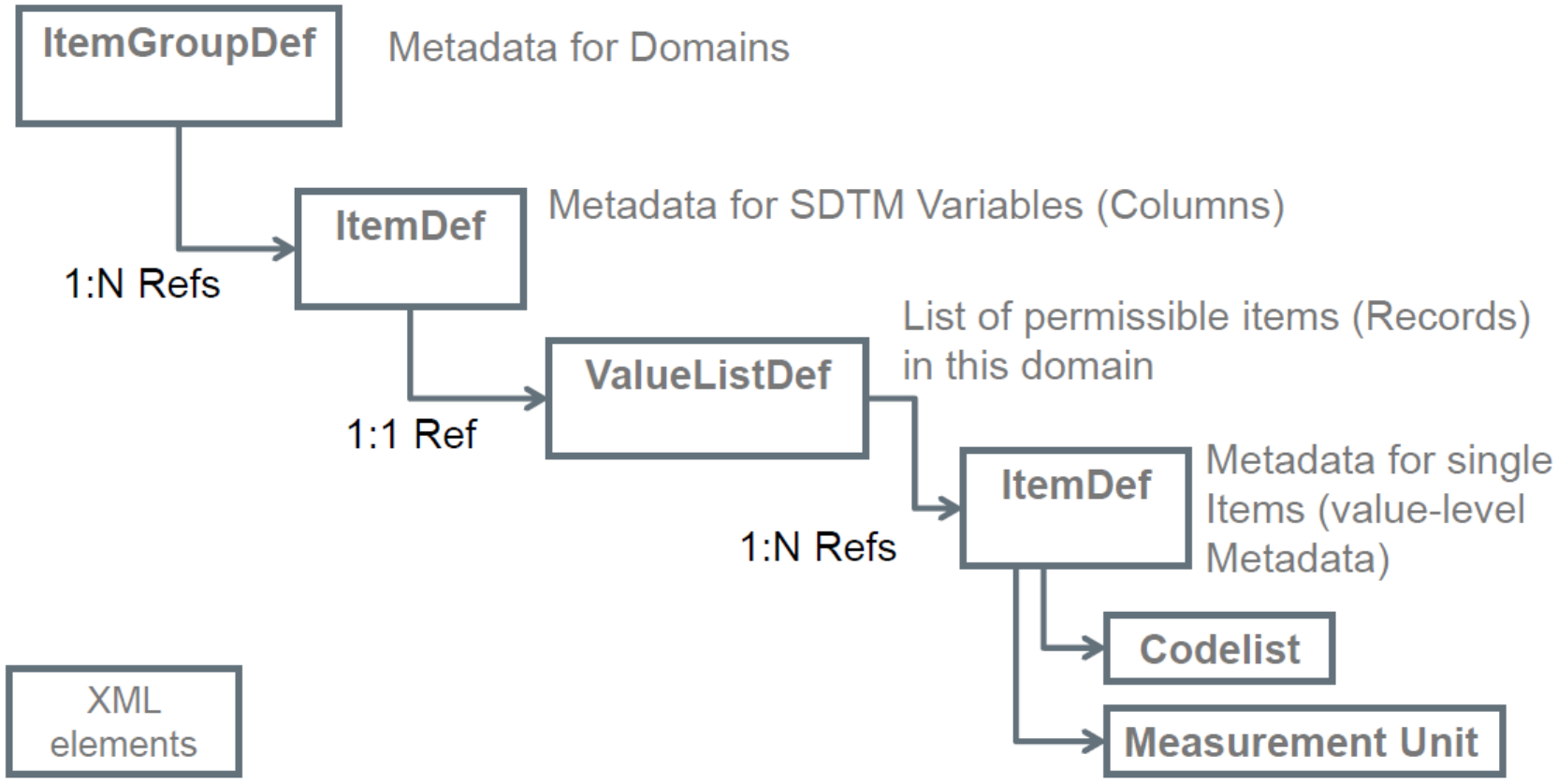
- **Metadata**
  - a description of the study data in a structured xml format designed to be machine and human readable
  - For a reasonable size SDTM grouping of datasets the define.xml may have up to 10,000 lines of xml text



What is Define.xml? – 2

- **The Define contains the complete set of data definitions for:**
  - the meaning / scientific concept
  - ID / Label / short name
  - (Long) Name / description
  - References to external classification systems (e.g., MedDRA)
  - the type of data (integer, float, text, date, time)
  - the maximum length of the data value
  - the possible / permissible units of measure
  - permissible discrete answers (codelists)

Structure of Define.xml



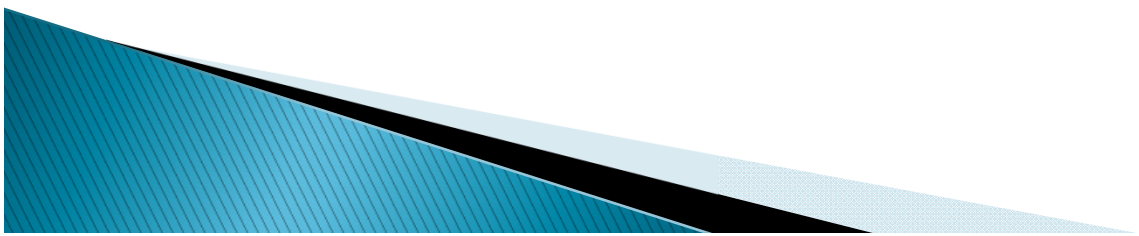
Raw XML list of Define.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ODM FileOID="2" FileType="Snapshot" CreationDateTime="2013-03-13T20:41:31-07:00" Originator="Formedix" SourceSystem="Origin Submission Modeller" SourceSystemVersion="2.10.2" ODMVersion="1.2" xmlns:fdx="
http://www.formedix.com/ns/origin/2" xsi:schemaLocation="http://www.cdisc.org/ns/odm/v1.2 define1-0-0.xsd" xmlns:def="http://www.cdisc.org/ns/def/v1.0" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="
http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.cdisc.org/ns/odm/v1.2">
  <Study OID="AALL07P2">
    <GlobalVariables>
      <StudyName>Erwinaze</StudyName>
      <StudyDescription>Erwinaze</StudyDescription>
      <ProtocolName>ALL07P2</ProtocolName>
    </GlobalVariables>
    <MetaDataVersion OID="4" Name="Erwinaze" Description="Erwinaze" def:DefineVersion="1.0.0" def:StandardName="CDISC SDTM" def:StandardVersion="3.1.2">
      <def:AnnotatedCRF>
        <def:DocumentRef leafID="blankcrf.pdf"/>
      </def:AnnotatedCRF>
      <def:leaf xlink:href="blankcrf.pdf" ID="blankcrf.pdf">
        <def:title>blankcrf.pdf</def:title>
      </def:leaf>
      <def:ValueListDef OID="ValueList.DA.DACAT">
        <ItemRef ItemOID="DA.DACAT.STUDY MEDICATION" Mandatory="No" OrderNumber="1"/>
      </def:ValueListDef>
      <def:ValueListDef OID="ValueList1.DA.DACAT.STUDY MEDICATION.DATESTCD">
        <ItemRef ItemOID="DA.DACAT.STUDY MEDICATION.DATESTCD.TREAT" Mandatory="No" OrderNumber="1"/>
      </def:ValueListDef>
      <def:ValueListDef OID="ValueList1.IE.IECAT.EXCLUSION.IETESTCD">
        <ItemRef ItemOID="IE.IECAT.EXCLUSION.IETESTCD.EX01" Mandatory="Yes" OrderNumber="1"/>
        <ItemRef ItemOID="IE.IECAT.EXCLUSION.IETESTCD.EX02" Mandatory="Yes" OrderNumber="2"/>
        <ItemRef ItemOID="IE.IECAT.EXCLUSION.IETESTCD.EX03" Mandatory="Yes" OrderNumber="3"/>
        <ItemRef ItemOID="IE.IECAT.EXCLUSION.IETESTCD.EX04" Mandatory="Yes" OrderNumber="4"/>
      </def:ValueListDef>
      <def:ValueListDef OID="ValueList1.IE.IECAT.INCLUSION.IETESTCD">
        <ItemRef ItemOID="IE.IECAT.INCLUSION.IETESTCD.IN01" Mandatory="Yes" OrderNumber="2"/>
        <ItemRef ItemOID="IE.IECAT.INCLUSION.IETESTCD.IN02" Mandatory="Yes" OrderNumber="3"/>
        <ItemRef ItemOID="IE.IECAT.INCLUSION.IETESTCD.IN03" Mandatory="Yes" OrderNumber="4"/>
        <ItemRef ItemOID="IE.IECAT.INCLUSION.IETESTCD.IN04" Mandatory="Yes" OrderNumber="5"/>
      </def:ValueListDef>
      <def:ValueListDef OID="ValueList.LB.LBCAT">
        <ItemRef ItemOID="LB.LBCAT.CHEMISTRY" Mandatory="No" OrderNumber="4"/>
```

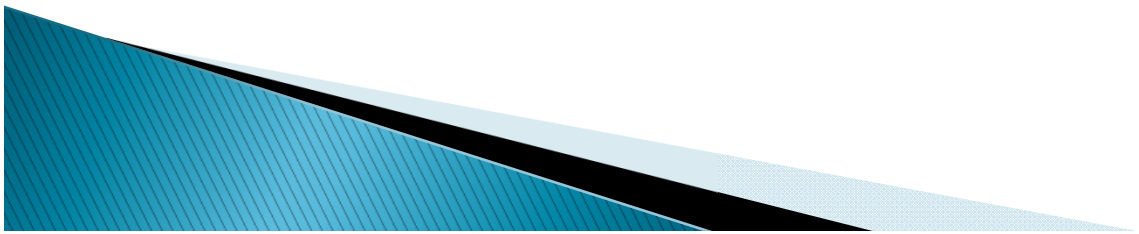
Before Validation Tools

- In the beginning…
  - Manual work
  - Custom programs / applications
  - Spreadsheet specifications to check and re-check

- There were no regulatory standards for validation until Nov 2014
  - OpenCDISC started with Janus Rules and WebSDM conformance

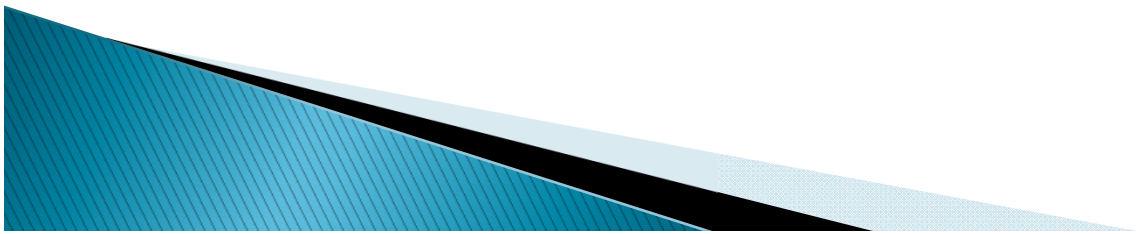


- Using Automated Validation Tools



Benefit of Tools

- Decreases time and increases accuracy
- Allows you to accomplish the impossible
  - Define.xml is metadata file that is hierarchical and based on ODM to support human and machine readability
  - A reasonable size study define.xml up to 10,000 lines of xml statements
- May tell you some things about your data you did not know



Limitations of Tools

- Extensible codelists create false errors
- Legacy studies have non-conformant terminology
- False positive results
  - Tests without units
- Different tools yield different results
- There is NO substitute for knowing your data structure and content
- There is NO substitute for consultation with Regulatory agencies

Interpreting the Validation Results – 1

What automation can for you

- dataset structure – variable names / labels
- data integrity
  - checks reference to DM subjects
  - presence of baseline flags
  - dates after disposition
  - Results units consistency
  - Terminology checks
- referential integrity
  - Start date before End date
  - Disposition references – sometimes

Interpreting the Validation Results – 2

What you still have to do

- dataset structure
  - do you have the right variables per spec?
  - custom domains
- data integrity
  - baseline flags – does not see 2 per subject?
  - Review Terminology flags, If a codelist is expandable – is it correct?
- data validation more content focused
  - right subjects
  - right dates
  - right codelists
  - right test codes

Interpreting the Validation Results – 3

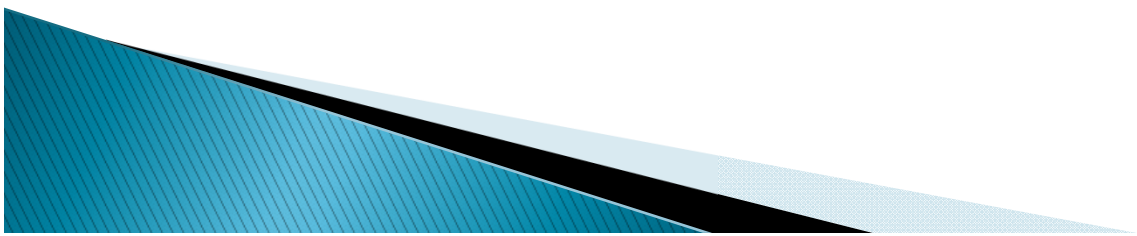
- False positive results may occur
  - Lab tests without units will generate errors e.g. Urine pH = 5, specific gravity 1.012
  - Terminology
    - Extensible codelists with non matches in the terminology file where the sponsor has added codes / values not present
      - E.g. Oxygen Saturation as O2SAT VSTESTCD
  - Some program bugs may generate false error messages
    - Report these!

Suggested Process for OpenCDISC Validator

- Review Setup / Parameters
  - Confirm SDTM Version
  - Confirm MedDRA version
  - Confirm Terminology version
  - Include the define.xml file if present
  - Set report parameters
    - Study Name / Number / Dates / other text / Excel message limit
- Run validation
- Review Error report
  - Update Issue Summary tab with Comments
  - Refer to details tab as needed
  - Identify / report any new bugs
  - Consider submitting an update to terminology team
- If final dataset for submission then include in reviewers guide data conformance section with explanations

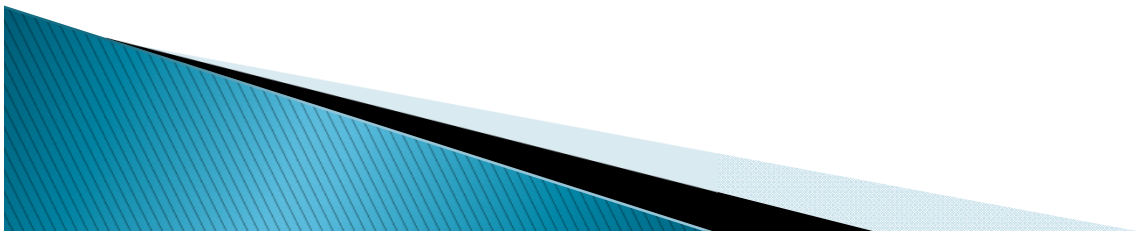
Lessons Learned – 1

- Focus on content more!!!
  - You know your studies, the data collected and analyzed better than anyone else
- Older studies more likely to be non-conformant
  - If the data was collected using non-standard terminology or older efficacy measures know the strengths and weaknesses of what you have



Lessons Learned – 2

- Understand the exceptions and their meaning
  - False positives
  - Non standard terminology
- Involve study team experts to understand and explain complicated clinical questions that are contained in
  - Statistical Analysis Plans
  - CRF Data and other source data



Lessons Learned – 3

- Provide documentation to explain what is known
  - Reviewers Guide
    - Hard codes
    - Non Conforming data
- Fix what is possible
- Be prepared to answer questions about legacy data

Thank you!

David Borbas RN, MIS
Senior Director, Data Management
Jazz Pharmaceuticals

David.Borbas@jazzpharma.com
P. 650-496-2637