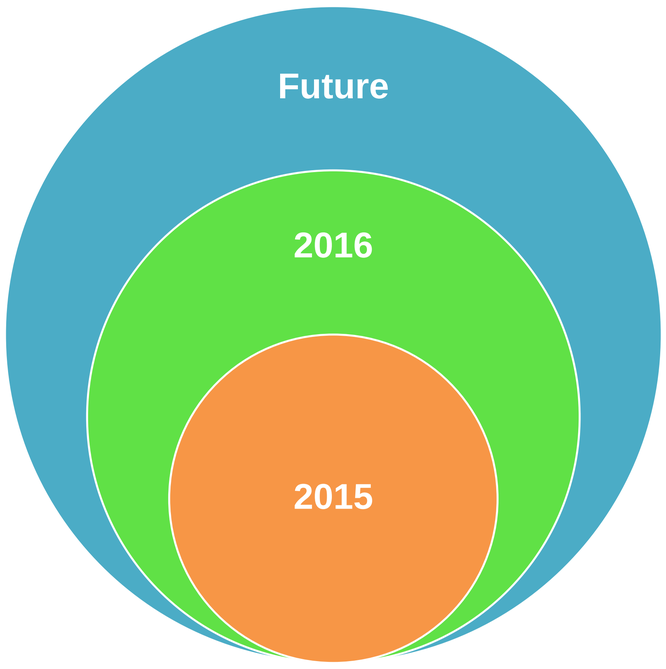
Common Errors in Loading SDTM Data to The Janus Clinical Trials Repository

Crystal Allard[1], John Slattery[2], Harsh Marwaha[3], Kathryn Matto[2], Lilliam Rosario[1]

1 Office of Computational Science, 2 IBM, 3 Ekagra

The Janus Clinical Trials Repository (CTR) is a data warehouse application that supports the automated extraction, transformation, loading, management, and integration of clinical trials data to support regulatory review and analysis needs. Developed in collaboration with the National Cancer Institute (NCI), the Janus CTR is a standards-based repository of subject-level clinical trial data that will facilitate the regulatory review of drug and biologic therapeutic products. The Janus CTR is designed to receive data in CDISC SDTM format and con support other study data exchange standards, such as CDISC SEND, HL7 v3, FHIR, and RDF. The Office of Computational Science (OCS) has been loading data to the Clinical Trials Repository since November 2014.

We tracked and inventoried all errors encountered in the loading of SDTM datasets for 180 clinical studies to date

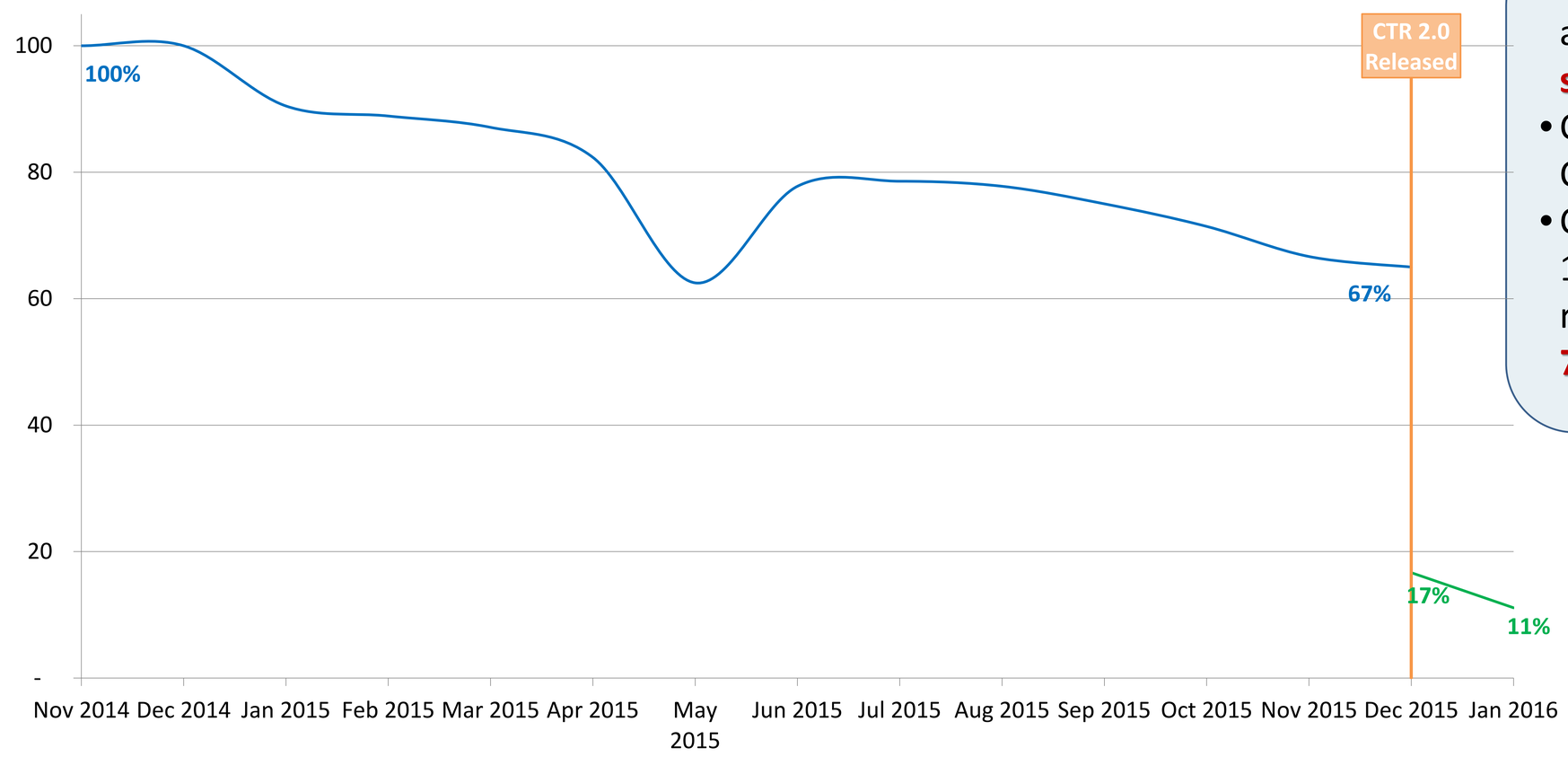


Future: Load all studies with SDTM datasets for all applications received at FDA

2016: Load all studies for applications that receive JumpStart Service in 2016. This includes the selected studies used during the JumpStart Service as well as any other studies within the application that contain SDTM datasets.

2015: Load selected studies for applications that receive JumpStart Service in 2014 or 2015. This includes only the selected studies used during the JumpStart Service. There is a maximum for 4 studies per jumpstart service.

CTR First Load Failure Rate by Month



- Could load on 1st attempt: **22% (39 studies)**
- Could not load to CTR: **4% (8 studies)**
- Could not load on 1st attempt and required curation: **74% (133 studies)**

86 Studies (55%) had major blocking issues	28 Studies (18%) were missing TS domain files	17 Studies (11%) had misc. loading issues	8 Studies (4%) did not load to Janus CTR
• 65 studies (43%) had issues within the Define files • 76 studies (49%) had issues within the TS domain data	• Had to create the TS domain file in order to load the studies	• Issues within other domain files that prevented the initial load or software problems that prevented the initial load	• 4 studies (2%) are missing define files • 4 studies (2%) have significant data issues and we cannot fix them without more information from sponsor/sponsor resubmission