

# The ISOcat Data Category Registry Standardization Process



STANDARDIZING DATA CATEGORIES  
&  
DATA CATEGORY SPECIFICATIONS



SUE ELLEN WRIGHT  
METADATA TDG WEBINAR  
2011-05-16

# Standardization in the DCR



- The DCR comprises a global collection of data category specifications used in a wide range of language resources.
- Individual data categories (DCs) are assigned to one or more Thematic Domain Groups (TDGs).
- The task of the Metadata TDG is to standardize the Individual DCs in the Metadata Data Category Selection.

# Two-Step Standardization Process



- Standardization in the DCR follows a two-step process.
  - Evaluation phase conducted by subject-area experts making up the *Thematic Domain Group* under the management of the *TDG Chair* (Peter Wittenburg for the *Metadata TDG*)
  - Validation phase conducted by the TC37-wide DCR Board, made up of experts nominated by the P-member bodies of ISO TC 37
  - Steps mirror standard ISO practice (Committee Draft [CD] followed by a Draft International Standard [DIS])

# Standardizing DCs



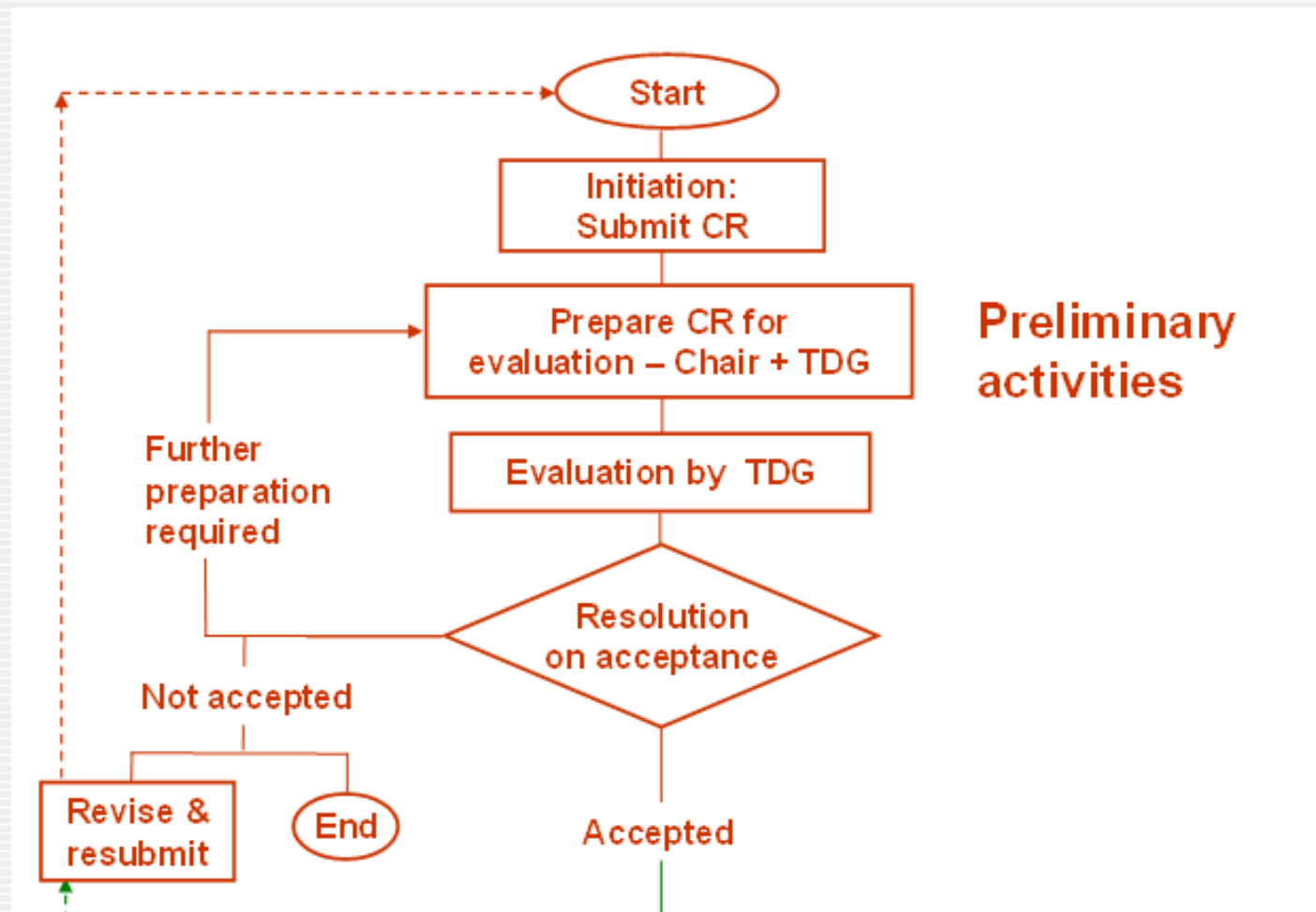
- Individual data categories are discussed and standardized.
- Important standardization elements include:
  - Identifier
  - Data category name
  - Data category type (container, complex, simple)
  - Data category definition
  - Conceptual domain
    - ✦ Can vary for same DC used in different TDGs

# Standardizing DCs

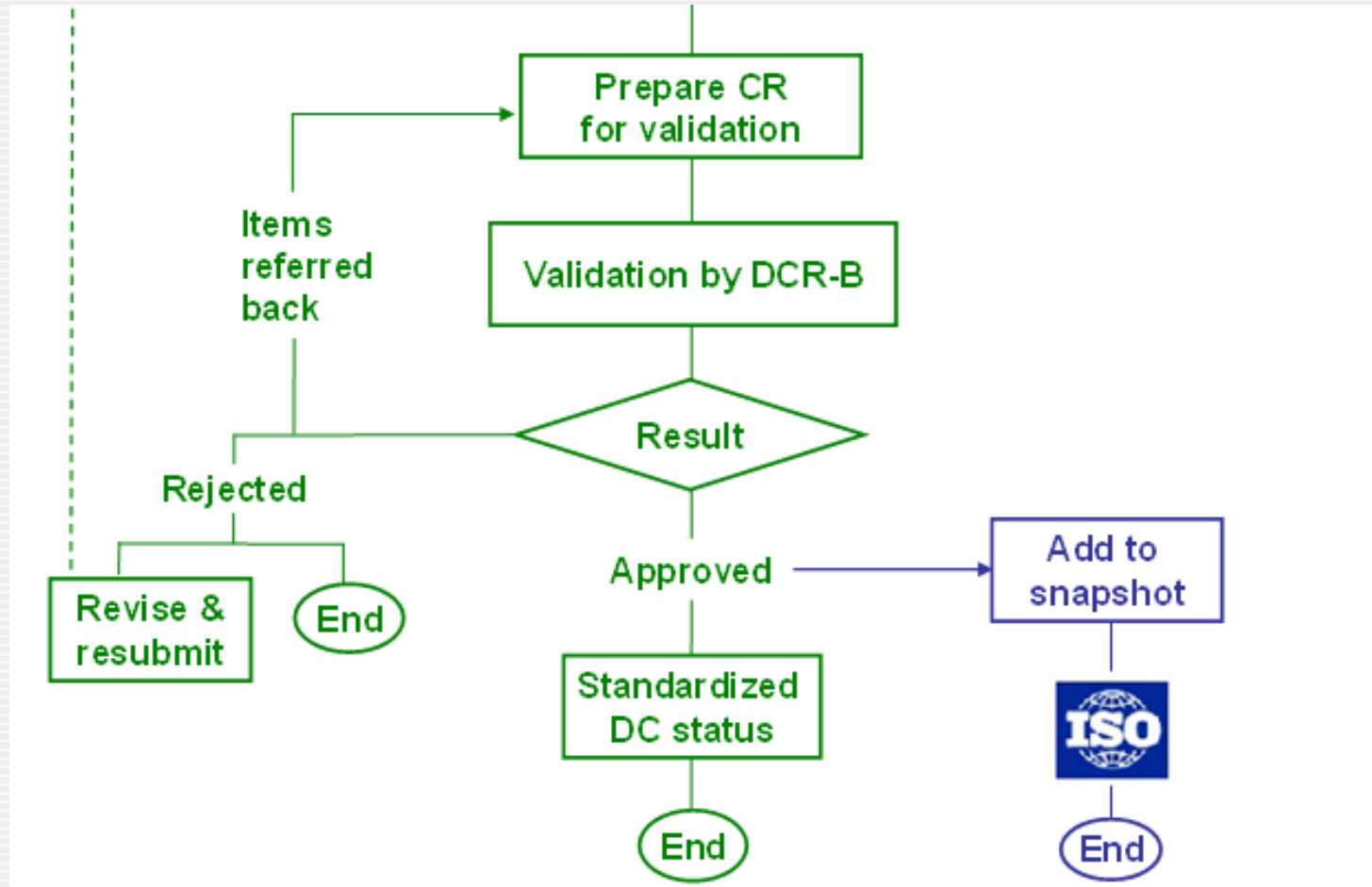


- **Excluded from standardization**
  - Optional translations in other Language Sections
  - Changes may take place during the standardization process, (e.g., in the data category name or in definitions, examples, explanations, notes), which will mean that language sections will have to be updated once the standardization process is complete.
  - Additional translations can be added at any time.

# TDG Evaluation Phase



# DCRB Validation Phase



# Standardized DCs



- Frozen after standardization
- Changes require submission of Change Request



# Standardizing DCSs



- Unresolved issue: Once all DCs in a DCS have been standardized it may or may not be desirable to standardize the whole set.
- If a DCS is used as a functional unit in the framework of an given annotation scheme (like TBX or LMF), it may be desirable to standardize the DCS in order to ensure interoperability.
- If the set is more open-ended and it is possible to pick and choose options depending on pragmatic needs, it may be desirable to leave the set open and not standardize at the DCS level.