

# Rainbow Extractor and Merger for XLIFF 2

FEISGILTT – Berlin – June 2015

Yves Savourel (ENLASO)

# Rainbow – some context

- Rainbow: A simple GUI tools to run Okapi processes
- The Okapi Framework: Many components performing localization-related tasks
  - One component extracts translatable data from various file formats into the Okapi internal object model
  - Another component creates various Translation Kit packages
  - Another is to merge the translated text back into the original format
- Various translation-kits formats available
- One of them is XLIFF 2 documents

# XLIFF 2 extractor agent

- Extraction: “The process of encoding localizable content from a native content or User Interface format as XLIFF payload, so that localizable parts of the content in the source language are available for Translation into the target language along with the necessary context information.”
- Responsible for creating the structure of the document, filling in the source content, setting permissions, etc.
- Does not necessarily perform segmentation, MT matches, etc.

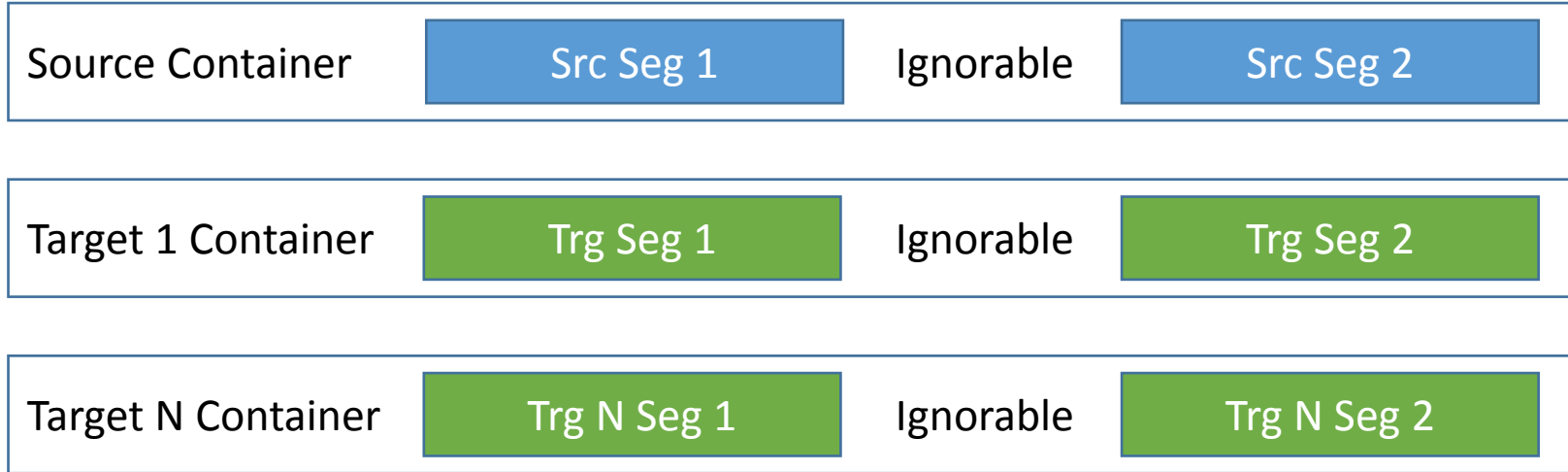
# XLIFF 2 merger agent

- Merging: “The process of importing XLIFF payload back to the originating native format, based on the full knowledge of the Extraction mechanism, so that the localized content or User Interface strings replace the source language in the native format.”
- Put back the modified content into its original format.
- Should be able to handle any allowed modifications to the original document (e.g. segmentation, annotations, etc.)

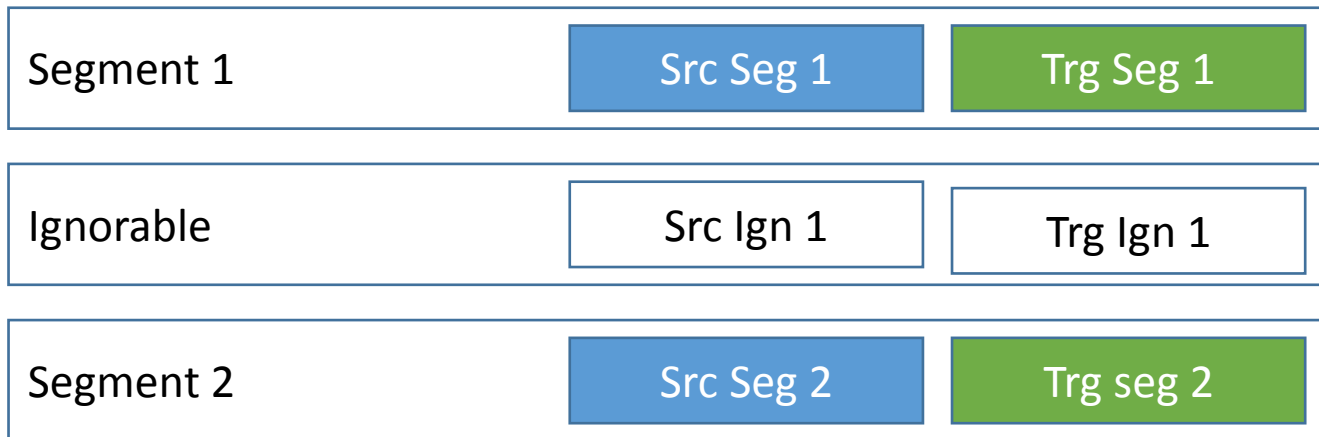
# Challenges – structure differences

- Okapi:
  - Source container
    - List of source segments
  - Target container (one or more)
    - List of target segments
- XLIFF 2:
  - List of segments or ignorable, in each segment/ignorable:
    - Source
    - Target

**Okapi**  
Text Unit



**XLIFF 2**  
Unit



# Challenges – unknown information

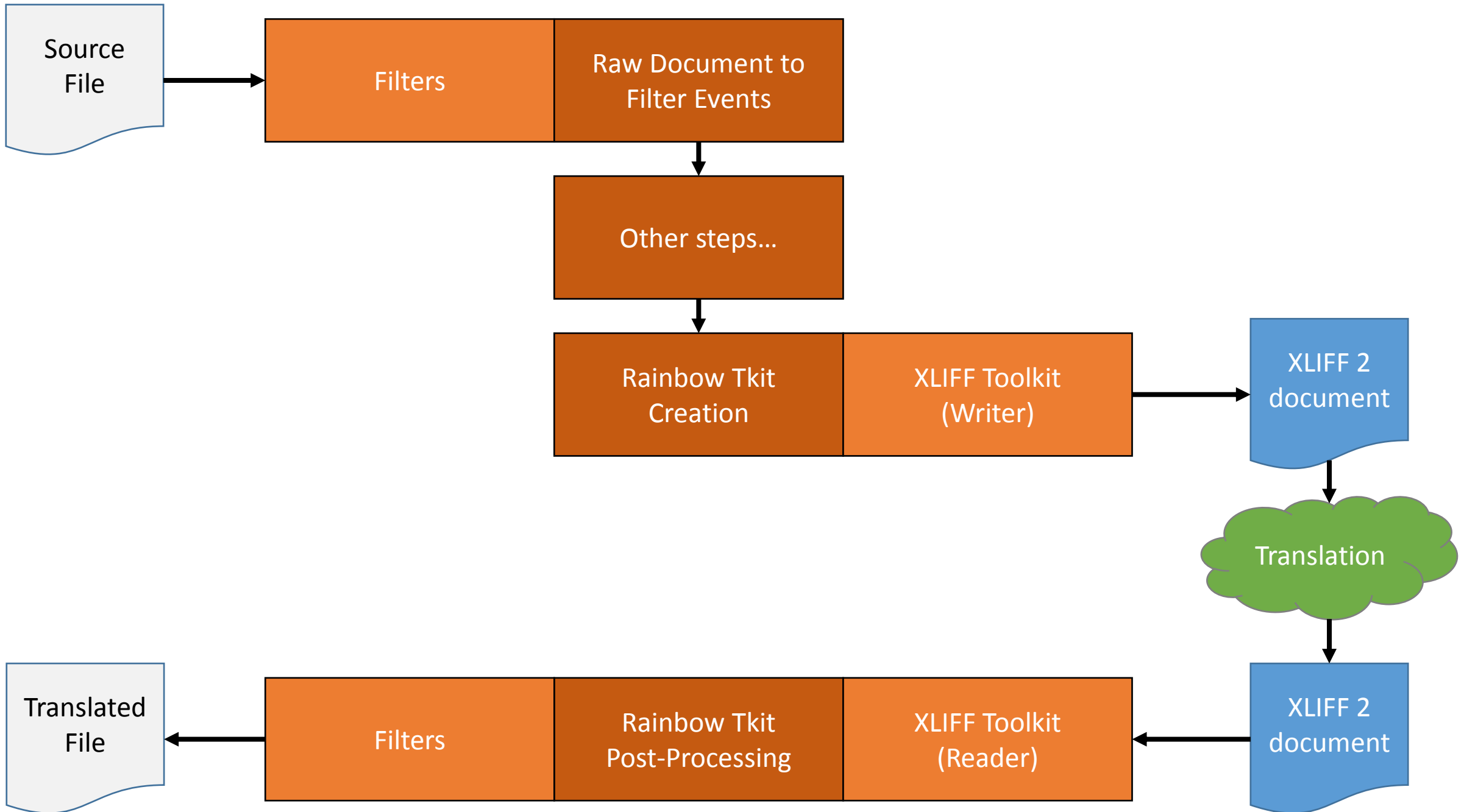
- Not all filters provide metadata such as whether an inline code can be copied, can be deleted, etc.
- Translation candidates may not have always all information such as suitability or quality.

→ Need to choose “good enough” default in some cases

# Okapi XLIFF Toolkit

- A library for creating, writing and reading XLIFF 2 documents
- Supports: the **Core** (obviously) and the modules: **Translation Candidates, Glossary, Metadata** and **Validation**
- Generic support for all other modules and extensions
- Download:  
[https://bintray.com/okapi/Distribution/Okapi XLIFF Toolkit](https://bintray.com/okapi/Distribution/Okapi%20XLIFF%20Toolkit)
- Maven artifacts:  
<https://repository-okapi-xliff.forge.cloudbees.com/release>
- Getting Started guide:  
<https://bitbucket.org/okapiframework/xliff-toolkit/wiki>





# An example...

Using Rainbow to extract and merge a couple of files in different formats



# Useful links

- XLIFF 2 specification:  
<http://docs.oasis-open.org/xliff/xliff-core/v2.0/xliff-core-v2.0.html>
- Okapi Applications download (includes Rainbow):  
[https://bintray.com/okapi/Distribution/Okapi Applications](https://bintray.com/okapi/Distribution/Okapi%20Applications)
- Okapi XLIFF Toolkit home:  
<https://bitbucket.org/okapiframework/xliff-toolkit>