

# The CLARIN-NL Data Curation Service (DCS)



Nelleke Oostdijk, Henk van den Heuvel, Maaske Treurniet

CLS/CLST, Radboud University Nijmegen, The Netherlands  
 {n.oostdijk,h.vandenheuvel}@let.ru.nl



## DCS main activities

1. Curation of language resources from the Netherlands, esp. those presently held by individual researchers or research groups;
2. Assisting in the curation efforts of CLARIN-NL Data Centres (if and when such is desired);
3. Advising researchers who wish to undertake the curation of their resources themselves.

## Tasks and subtasks

### Task A. Identification and assessment

1. Identify candidate resources; collect info as to
  - a. the owner/producer
  - b. the type of resource
  - c. the licensing restrictions/conditions
  - d. the size
  - e. the format(s)
  - f. the metadata available
  - g. the nature of enrichment / annotations
2. Assess the desirability of curation
3. Assess the feasibility of successful curation

### Task B. Development of a curation plan

4. Evaluate the content objects and determine
  - a. what type and degree of format conversion or other preservation actions should be applied
  - b. the appropriate metadata needed for each object type and how it is associated with the objects
5. Estimate cost and lead time
6. Arrange for the necessary expertise to be available

### Task C. Curation

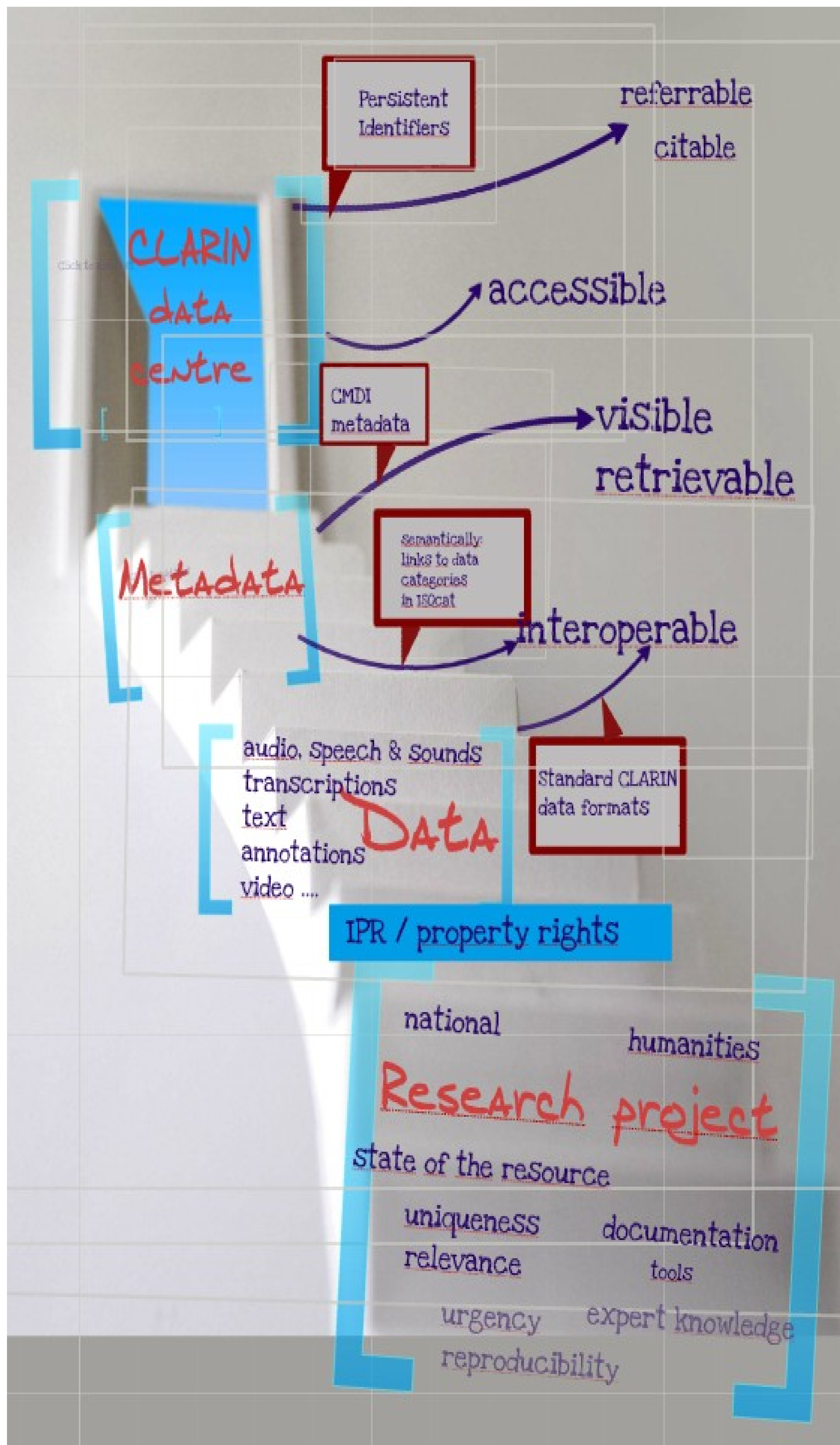
7. Digitize data (minor)
8. Convert to a CLARIN preferred format
9. Assign appropriate CMDI metadata
10. Ensure semantic interoperability
11. Provide documentation

### Task D. Validation

12. Validate curated resource

### Task E. Archiving

13. Transfer to CLARIN Data Centre for long-term storage and maintenance
14. Assign persistent identifier(s)
15. Provide access to content



## Examples of curated data

Dutch Bilingual Database  
 Dialect lexicons  
 Interviews with Veterans

## References

<http://www.clarin.nl/node/147>  
<http://www.ru.nl/letteren/datacuratieservice/>