CLARIN

Common Language Resources and
Technology Infrastructure

CLARIN and DOIs in the UK:
some issues for the Oxford Text Archive

Martin Wynne

Director of User Involvement, CLARIN ERIC

martin.wynne@it.ox.ac.uk

CLARIN Annual Conference

Thursday 23rd October 2014

Oxford e-Research Centre &

IT Services (formerly OUCS) &

Faculty of Linguistics, Philology and Phonetics,

University of Oxford

- There are national and institutional **policies** for DOIs in terms of:
  - Eligibility (Who can have one, for what resources?)
  - Procedures for allocation (Via which authority? How?)
  - Syntax of DOI naming (How much freedom?)
  - Payment (Who pays, how much?)

  (e.g. resources currently have to be in the Bodleian Library's institutional repository to be allocated a DOI by the University of Oxford, and the repository only collects outputs from Oxford academics)

- Decisions to be made about **granularity** of assignment (different requirements for citation and for automatic processing by language processing tools?)
- Resources in multiple repositories - lots of **copies**, lots of DOIs?
- **Bulk** assignment – what do you do if you have 4000 to register?
- But it's not all negative – there is much to gain to monitor citation and usage, and to be able to work with resources not curated by CLARIN

Format of DOIs

- ➢ Format : **XX.XXXX/bodleianNNNN.V?file=FFFF** [1]
  - ▪ **X =** DOI prefix (10.5287)
  - ▪ N = Accession code for the data package
  - ▪ V = Version number for the data package (N)
  - ▪ F = Filename (optional)
    - • Sample DOI: 10.5287/bodleianfdk5.3
    - • 10.5287/bodleianfdk5.3/Readme

- ➢ DOIs are defined to be case-insensitive

- ➢ Version number can be omitted

[1] The proposed syntax is based on Dryad model (see https://www.nescent.org/wg_dryad/DOI_Usage)