

LD4LT session

Metadata for Language Resources & Technologies

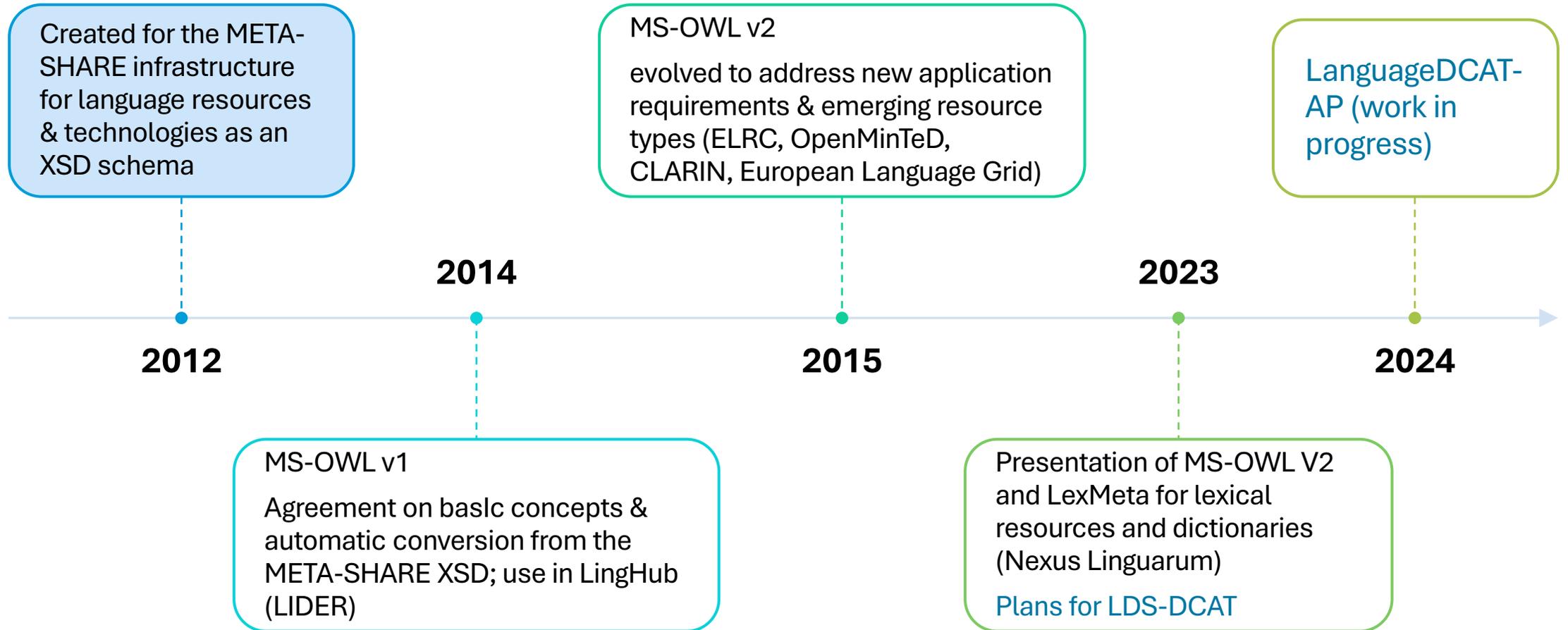
Penny Labropoulou

ILSP/Athena RC

Why?

- Need for cataloguing different types of datasets and data services in the wider Linguistic Data community
- Need for a common metadata schema for their description and identification
- Need for common encoding of usage rights across governmental, academic and commercial data sources

META-SHARE OWL (MS-OWL)



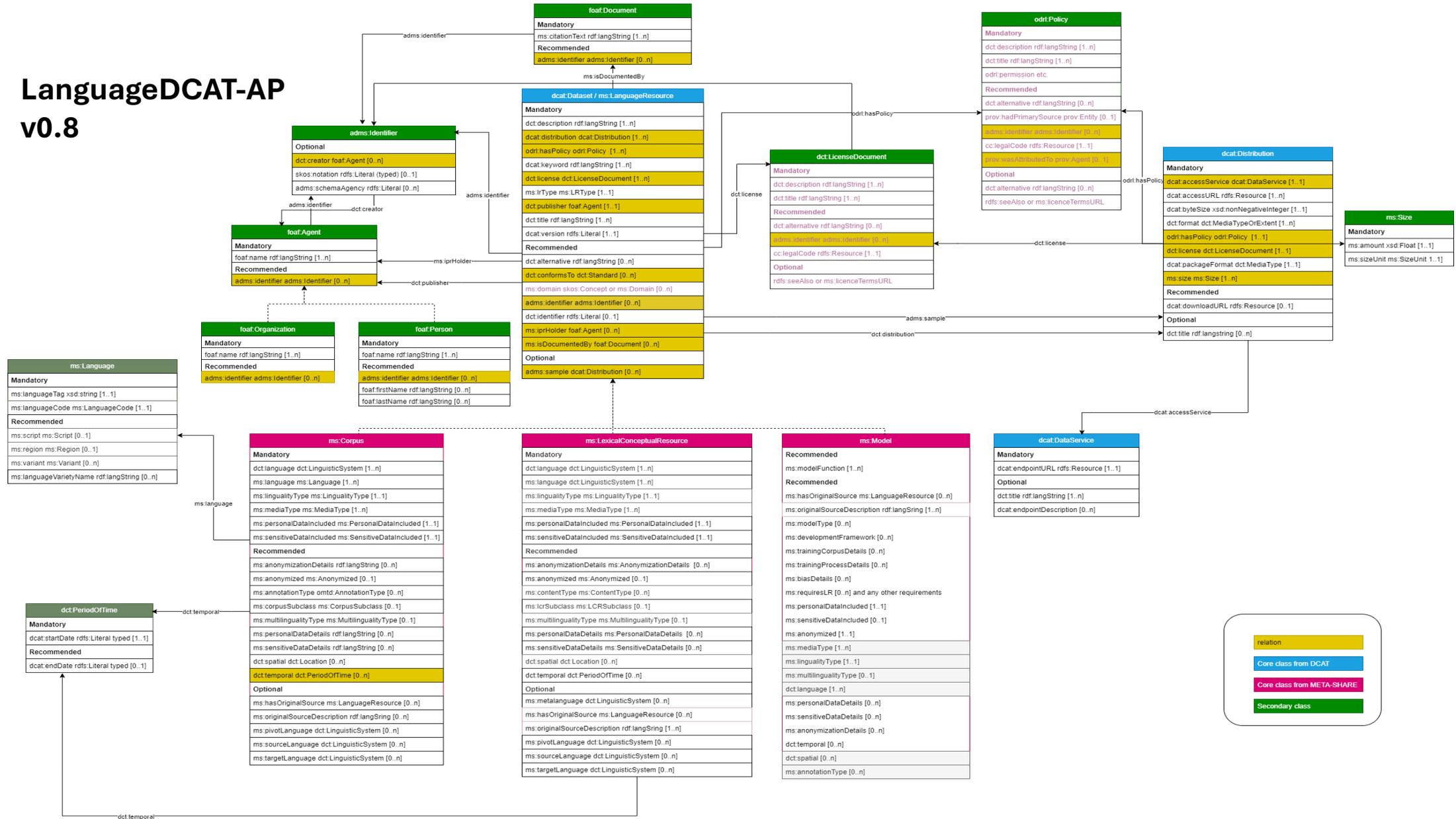
Why LanguageDCAT-AP?

- Why DCAT?
 - Widely adopted standard for cataloguing data sets and data services
 - LD4LT Wiki on discussions for [mapping META-SHARE with the DCAT model](#)
 - Emergence of Common European Data spaces & [European Language Data Space](#)
 - Semantic interoperability across data spaces: DCAT & DCAT application profiles for descriptions of data and data services & ODRL for the representation of policies ([DSSC blueprint](#))
- Why DCAT-AP?
 - [DCAT-AP v3](#): DCAT application profile for Catalogues containing Datasets and Data services descriptions in Europe, under the responsibility of the SEMIC group
 - it provides “improved definitions, usage notes and usage constraints such as cardinalities for properties and the usage of controlled vocabularies” (implemented as SHACL shapes)
 - widely adopted (domain-specific and national DCAT-AP customisations)

LanguageDCAT-AP: Main features

- "Reconciliating" requirements imposed by
 - technical specifications of Data Spaces (DCAT, ODRL, Data Space Protocol)
 - DCAT-AP specifications (e.g., EU controlled vocabularies)
 - the language data community needs and best practices
- Combining
 - DCAT for generic concepts & META-SHARE for language data specific concepts
 - Controlled vocabularies in SKOS format from DCAT-AP recommendations & from META-SHARE and OMTD-SHARE vocabularies

LanguageDCAT-AP v0.8



MS-OWL vs. LanguageDCAT-AP

- NOT a replacement
 - Cater for legacy data (e.g. META-SHARE, ELG, etc.)
 - MS-OWL addressing more focused/detailed needs depending on use cases and applications, e.g. detailed relations between language resources vs. `dct:isRelatedTo`
- Implemented as relations of classes/properties between the two models, e.g.
 - `ms:Corpus`, `ms:LanguageDescription` and `ms:LexicalConceptualResource` as subclasses of `dcat:Dataset` and `ms:ToolService` as `dcat:DataService`
 - `ms:resourceShortName` as subproperty of `dct:alternative`
- Improvements from LanguageDCAT-AP feeding MS-OWL

Where are we?

- Delays in setting up the LD4LT activities
 - prior engagements of the responsible chair
- Current status (by end of June)
 - Draft documentation of MS-OWL & LanguageDCAT-AP with examples
 - Pre-final OWL files
 - Pre-final SKOS vocabularies

Where do we go and how?

- Share documentation and get feedback from the community
- Enrichments/Improvements
 - Improve recommended properties and vocabularies (language, formats & data models, service functions, annotation types)
 - Missing properties (e.g. quality aspects)
 - Translation into other languages
 - Collecting new needs from community
- How?
 - Reviving LD4LT wiki and github
 - Initiating regular LD4LT calls – volunteers for helping their co-ordination?

Thank you!
Time for discussion