META≡NET

META-SHARE : the open exchange platform
Overview-Current State-Towards v3.0

Stelios Piperidis

Athena RC, Greece

spip@ilsp.gr

META≡FORUM 2012

A Strategy for Multilingual Europe
Brussels, Belgium, June 20/21, 2012

Introduction

META-NET

- Data is the crude oil of today's research technology development

- On all language technology related lists (corpora-list / linguist-list / mt-list, ..., requests
  - from English-X parallel corpora to syntactically annotated corpora
  - whether they are aligned, validated, ...
  - language identification in twitter streams

- Similar requests (domain-independent needs) articulated during all Vision Group meetings on the way to the SRA

- Data, need data, more data, big data, open linked data, ...

Introduction

META-NET

- However, only a portion of language resources is known / announced / shared / traded / …

- But…data collection, cleaning, annotation, curation and maintenance is a very costly business

- As evidence from other domains (e.g. biotechnology, geodata, earth sciences) shows data and tools become valuable through opening and sharing
  - Both for research and technology development
  - Evaluation
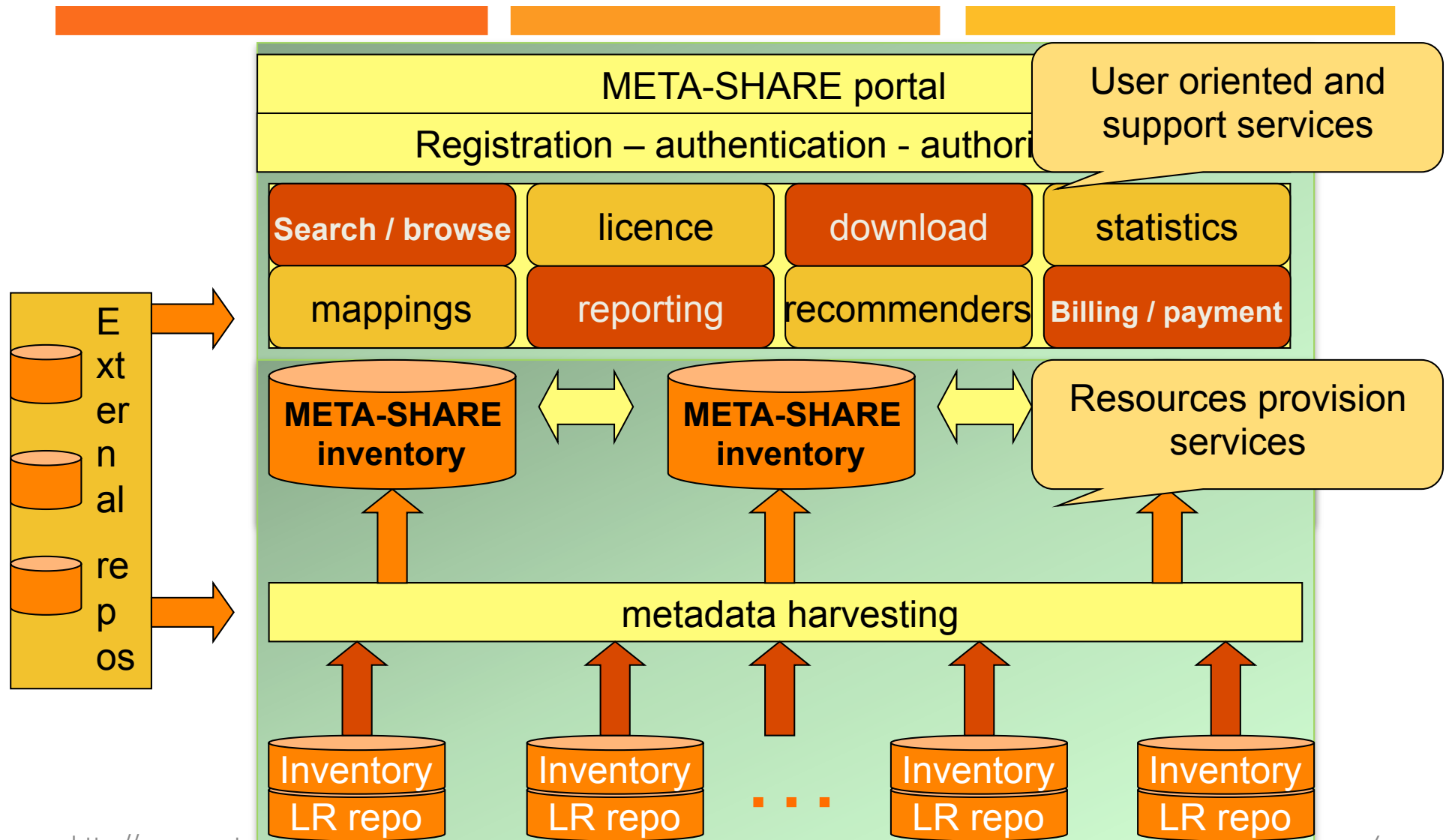 - Supporting innovative applications

META-SHARE rationale

- Language resources (data and tools) are dynamic living entities
  - they evolve over time in various dimensions (quantity, annotation levels, converted to a new format, addition of new languages)
  - they are usually the product of collaborative work
  - they may come with varying restrictions, …
- Need solutions that enable every **language resource provider**, at any granularity level (individual/lab/organisation), to
  - Create his **own store of LRs**
  - Describe, document and update it
  - **Link to a network** of other providers
  - **Keep track** of the use of his LRs, **trade** LRs
- Need solutions that enable every **language resource consumer** to
  - **Discover** what LRs suitable for his purposes exist
  - Get information about, **download or acquire** them

META-SHARE: what it is

- META-SHARE tries to match LR providers and consumers needs and expectations by enhancing **visibility, documentation, identification, availability, preservation** of language data and (basic language processing) tools

- It is an **open, non-monolithic, expandable, secure** exchange infrastructure for language data and tools for the Human Language Technologies domain

- A rather **long-term multidimensional endeavour** by which language resources can **boost research, technology and innovation** through **wide availability, pooling, openness and sharing**

META-SHARE architecture

META=NET

META-SHARE portal

Registration – authentication - authori...

User oriented and support services

Search / browse

licence

download

statistics

mappings

reporting

recommenders

Billing / payment

External repos

META-SHARE inventory

META-SHARE inventory

Resources provision services

metadata harvesting

Inventory LR repo

Inventory LR repo

. . . .

Inventory LR repo

Inventory LR repo

META-SHARE provider side

- ❑ META-SHARE is a **network of distributed repositories of LRs**
  - Local (organisation-based), and central repositories
  - Facilities for documenting, updating descriptions, storing/ linking LRs
  - Provider support services (forum, knowledge base)
  - Each repository maintains an inventory with all LRs MD, exports MD for harvesting
  - Harvested MD are stored in synchronised central servers

Add Distribution

Availability: available-unrestrictedUse
Specifies the availability status of the resource; restrictionsOfUse can be further used to indicate the specific terms of availability

Licences (Hide)
Licence: #1

Licence:
AGPL
ApacheLicence_V2.0
BSD-style
CC
The licence of use for the resource

Restrictions of use:
academic-nonCommercialUse
attribution
commercialUse
evaluationUse
Specifies the restrictions imposed by the type of the licence

Distribution access medium:
accessibleThroughInterface
bluRay
CD-ROM
downloadable
Specifies the medium (channel) used for delivery or providing access to the resource

Download location:
Add Another Field
Any url where the resource can be downloaded from

Metadata-based descriptions of LRs

Aim

- to support META-SHARE users (LRs providers and consumers) in all services provided
  - LR description (creation, storage and editing)
  - search and retrieval,
  - browse,
  - uploading & downloading resources
  - metadata harvesting/updating,
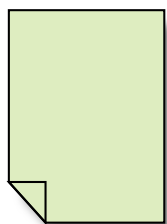 - monitoring of LRs and related objects, etc.

Metadata schema – ontology

META-NET

- entities described
  - core entity: the LR
  - satellite entities: related objects, e.g.
    - *actor*: *persons* and *organisations* involved, such as creators of resources, funders, distributors, etc.
    - *document*: reference documents, such as papers describing the resource, reports, tagset manuals, guidelines for LR production, etc.
    - *project:* projects that have funded the creation of an LR, or where an LR has been used, etc.
    - *licence:* used for the distribution of the LR

Ontology excerpt

META-NET

Entity

Language Resource | Licence | Project | Actor | Document

Corpus
Lexical / Conceptual Resource
Tool/Service
Language Description

Accompanies

Creator
Annotator
Distributor

Person
Organisation

Documentation
Annotation Manual
Validation Report
Standards Used
....

Core Entity   Satellite Entity
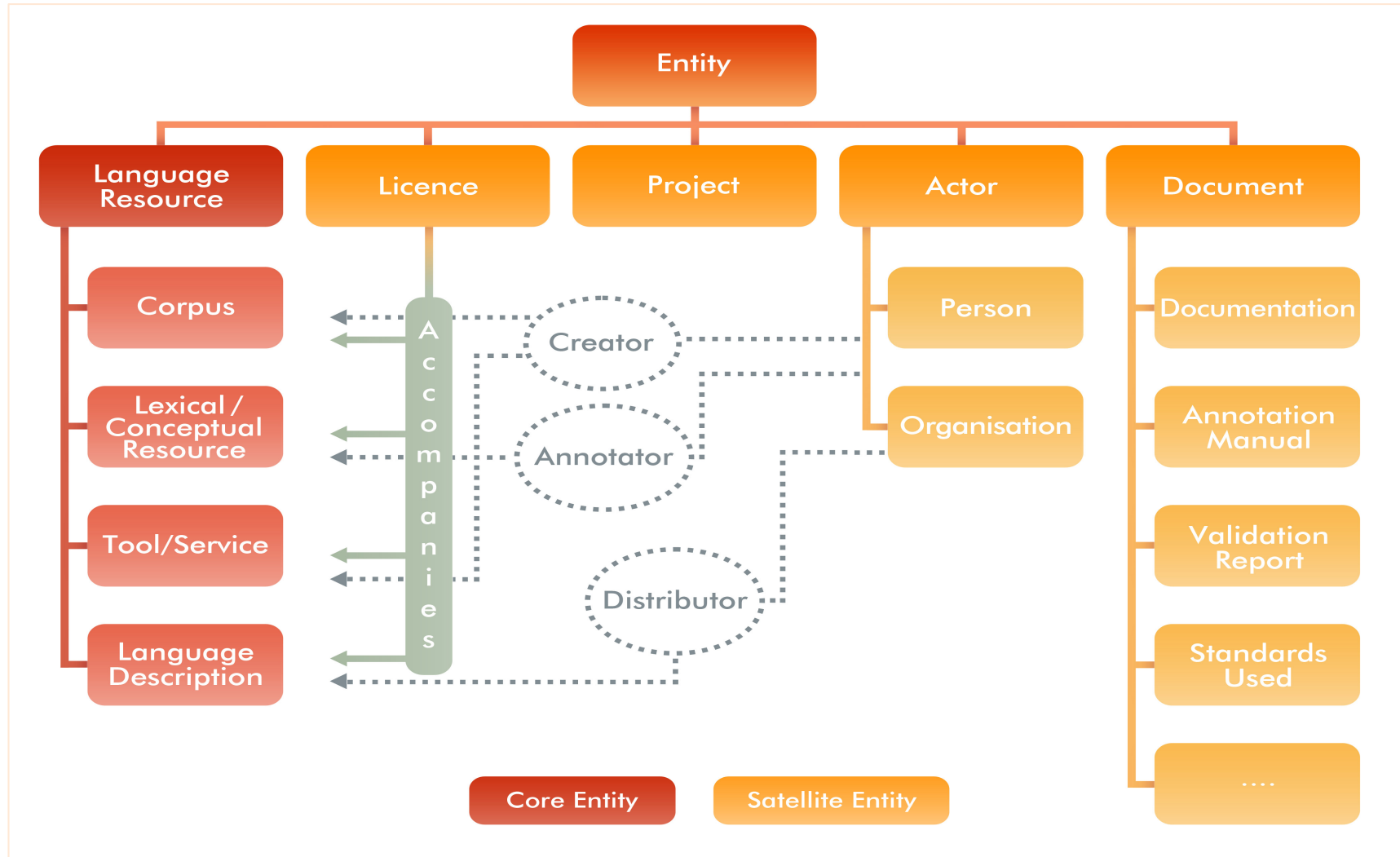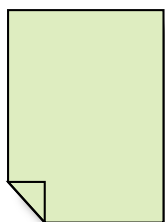
LRs typology (1)

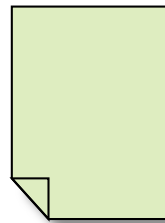- mediaType: text, audio, image, video

written corpora
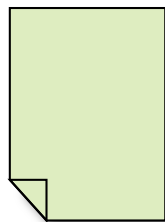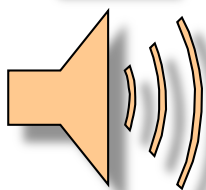
spoken corpora
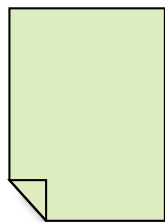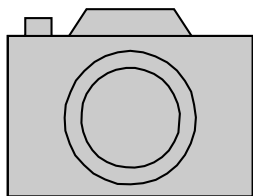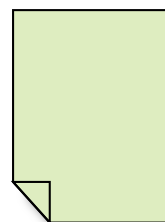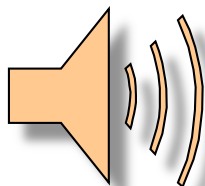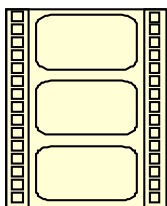
images (multimedia)
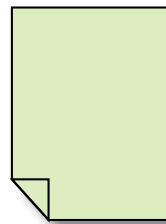
videos
(multimedia)

LRs typology (2)

META-NET
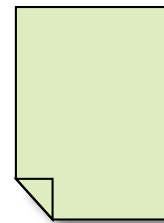
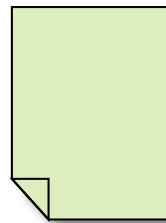- ❑ search for text
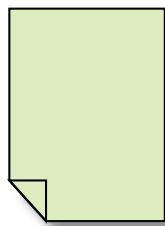
    written corpora

   spoken corpora

    images (multimedia)

    videos (multimedia)

META-SHARE user side

- Users (LR consumers) can
    - **search** the central inventory
    - **browse** using **multiple facets**
    - access the actual resources by visiting the respective repositories to get legally interoperable **licence(s)** to **download** and **use** them

    - get support through an **online user forum** and **helpdesks** dedicated to technical, metedata and legal issues
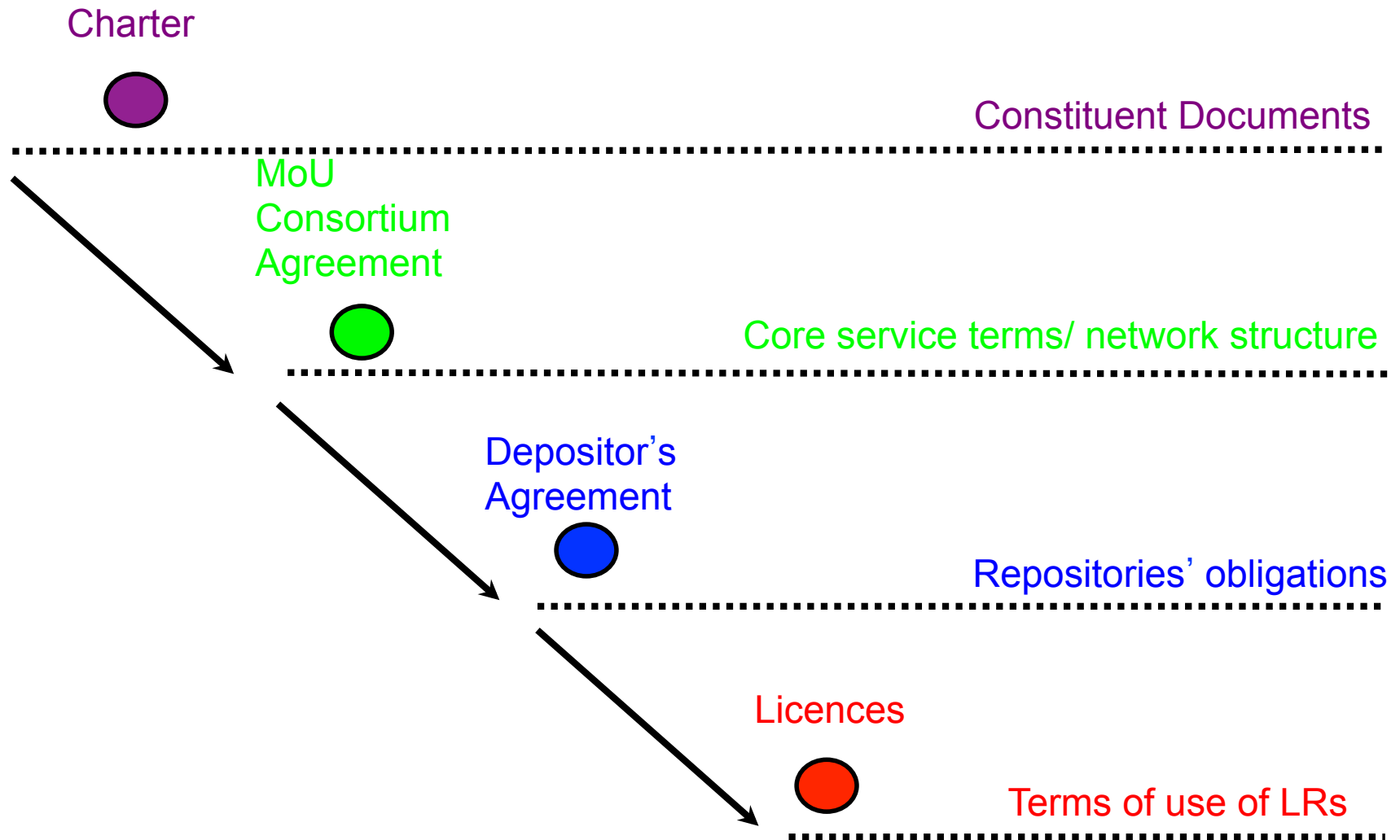   - access a **knowledge base**
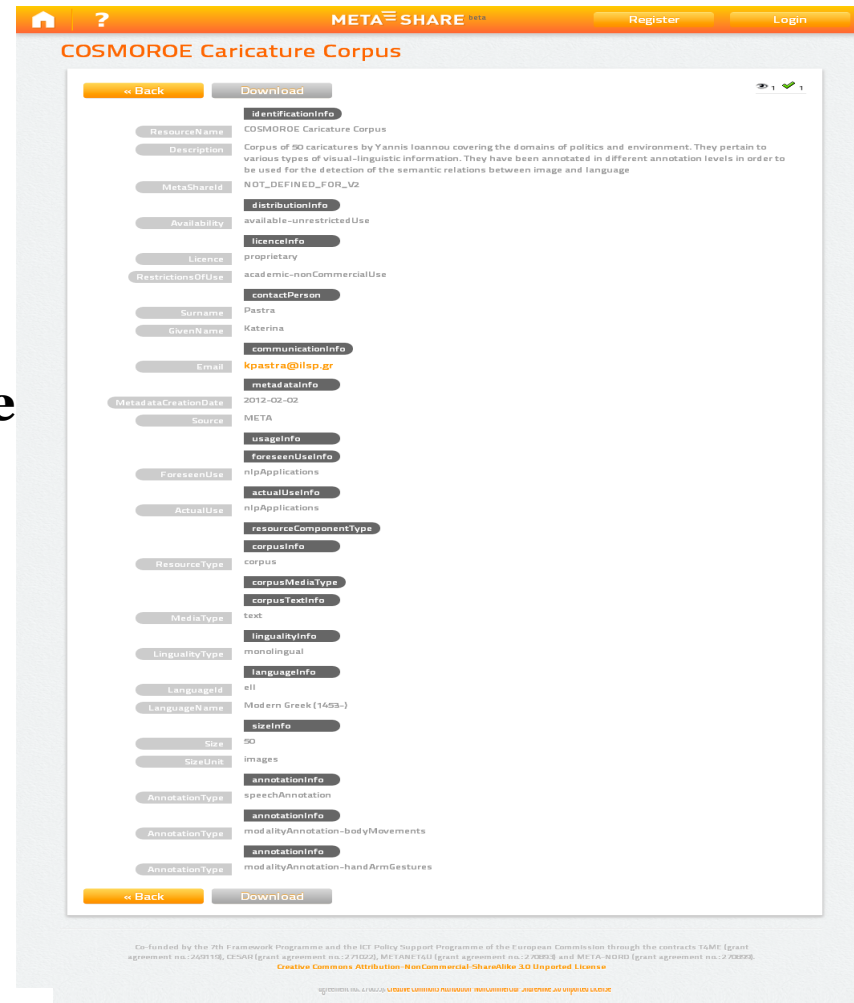
META-SHARE user support services

❑ Versions 2.0, 2.1 come with an online forum

❑ The purpose of the Forum is to :
- better manage the user support services
- monitor questions asked – answered - pending
- avoid repetition (both in asking and answering)
- help crystallise correct/mostly acceptable answers to user questions
- transform them into a useful wiki/knowledge base

Organisational and Legal Framework

META-NET

Charter

Constituent Documents

MoU
Consortium
Agreement

Core service terms/ network structure

Depositor's
Agreement

Repositories' obligations

Licences

Terms of use of LRs

Legal provisions

META⁼NET

- ❑ Language Resources Sharing Charter – high level principles

- ❑ Memorandum of Understanding – aka membership agreement

- ❑ Licensing templates and deposition agreements
  - ▪ Inclusive mix of open and openness inspired models
    - – **Creative Commons** licences (starting with Creative Commons Zero (CC-0) and all possible combinations along the CC differentiation of rights of use)
    - – **META-SHARE Commons** licences, fully developed CC-based licensing tool that allows META-SHARE members to make their resources available inside the network only
    - – **META-SHARE** "**No Redistribution**" licences, allowing use and exploitation of the Resources while permitting the LR Owner to have full control over the Resource distribution.
    - – Software tools and web services are either provided though one of the standard **Open Source** licenses or under a custom commercial license.

META-SHARE legal features

META=NET

❑ Rights based on type of use rather than type of user

❑ Differentiation along the following axes

- **Attribution or No Attribution**
- Open – share with everybody or within the network only
- **Redistribution vs No Redistribution**
- Commercial – non commercial
- **Derivative vs Non-Derivative**
- Share alike
 - Re-deposition of derivatives , as a soft norm in the membership agreement, to act as a driver for collaborative LR building

...share inside META-SHARE?

META-NET

|	BY	Commercial	Derivatives	SA
MSC BY	Y	Y	Y	N
MSC BYSA	Y	Y	Y	Y
MSC BY NC	Y	N	Y	N
MSC BY NC SA	Y	N	Y	Y
MSC BY ND	Y	Y	N	N
MSC BY NC ND	Y	N	N	N

...protect the original?

META NET

|	Commercial	Redistribution	Derivatives	Fee
MS Commercial NoRed FF	Y	N	Y	Y
MS Commercial NoRed	Y	N	Y	N
MS Commercial NoRed NoDer FF	Y	N	N	Y
MS Commercial NoRed NoDer	Y	N	N	N
MS NonCommercial NoRed NoDer FF	N	N	N	Y
MS NonCommercial NoRed NoDer	N	N	N	N
MS NonCommercial NoRed FF	N	N	Y	Y
MS NonCommercial NoRed	N	N	Y	N

Some remarks

META=NET
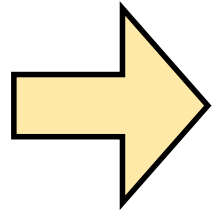
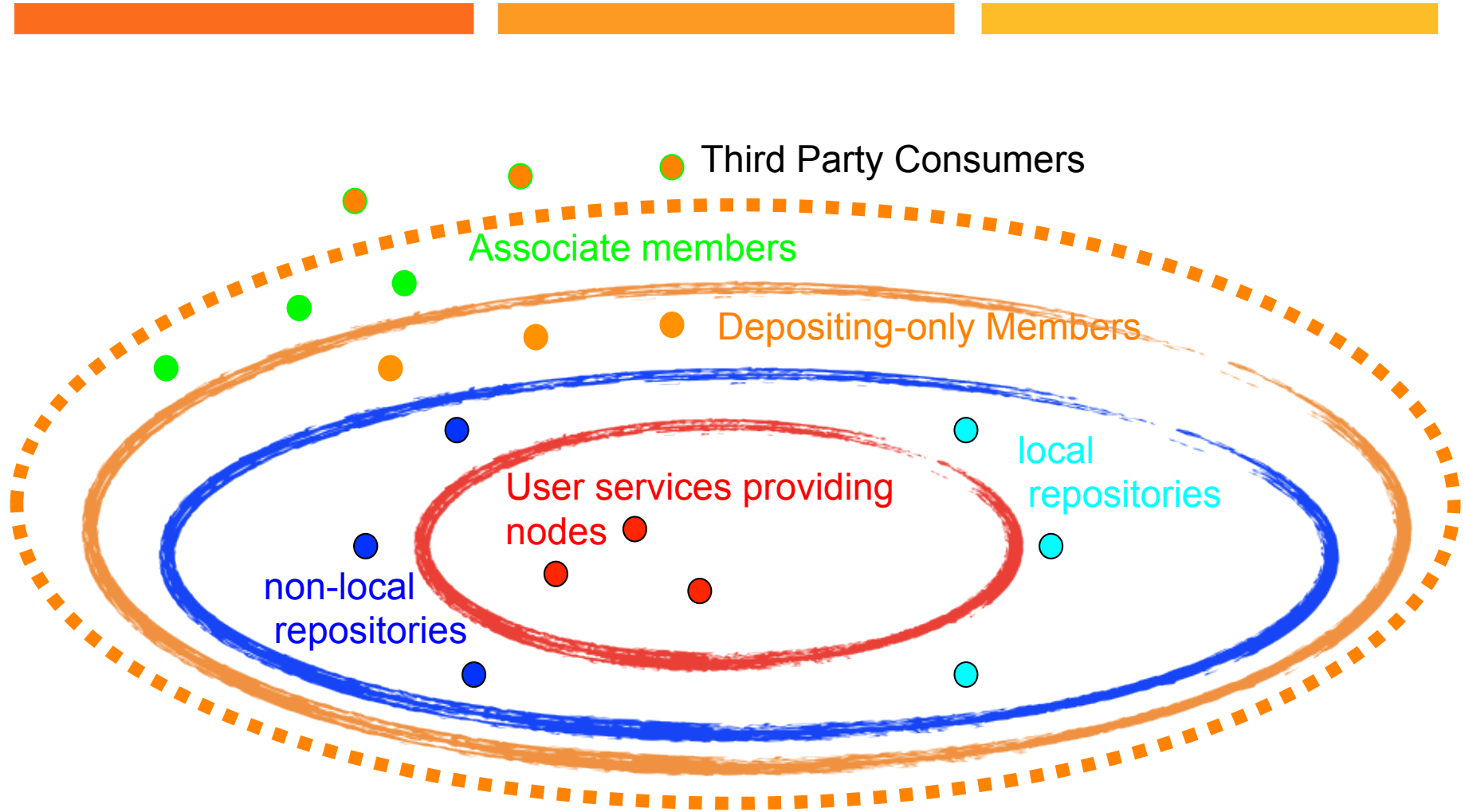→ clear before sharing or opening
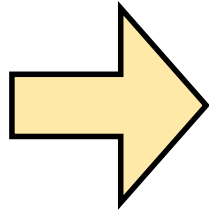
→ use standard licences
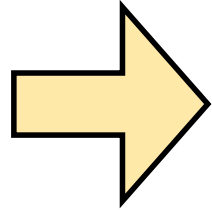
→ don't limit unless you need
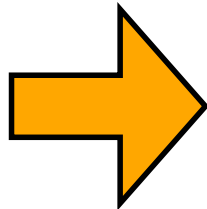
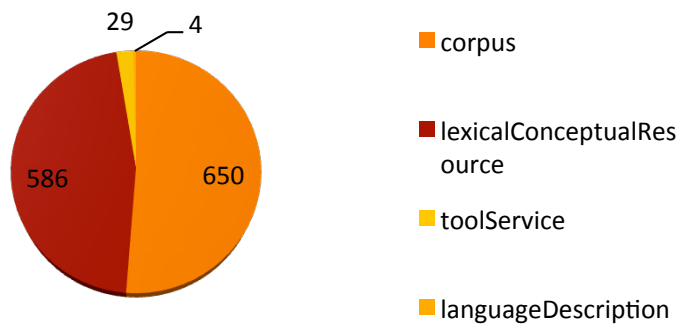→ use METASHARE services

Network structure

To sum up...

- META-SHARE software, open source, under a permissive licence (BSD), to set up a language resource repository

- Legal instruments catering for a range of uses

- Mapping services to big resource inventories

- Software-based services for both LR providers and LR consumers
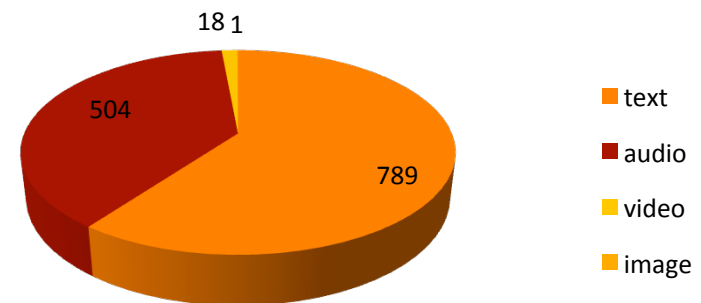
- User support services
  - User Forum
  - helpdesks

META-SHARE Repos v2.0

Resource types

29   4

- corpus
- lexicalConceptualResource
- toolService
- languageDescription

586   650

Media Type

18 1

504   789

- text
- audio
- video
- image

Distribution per Language

30 27
36 28 25 21
38   20
40
41   293   503
76
99   202   204   391

- English
- Spanish
- French
- German
- Italian
- Chinese
- Dutch
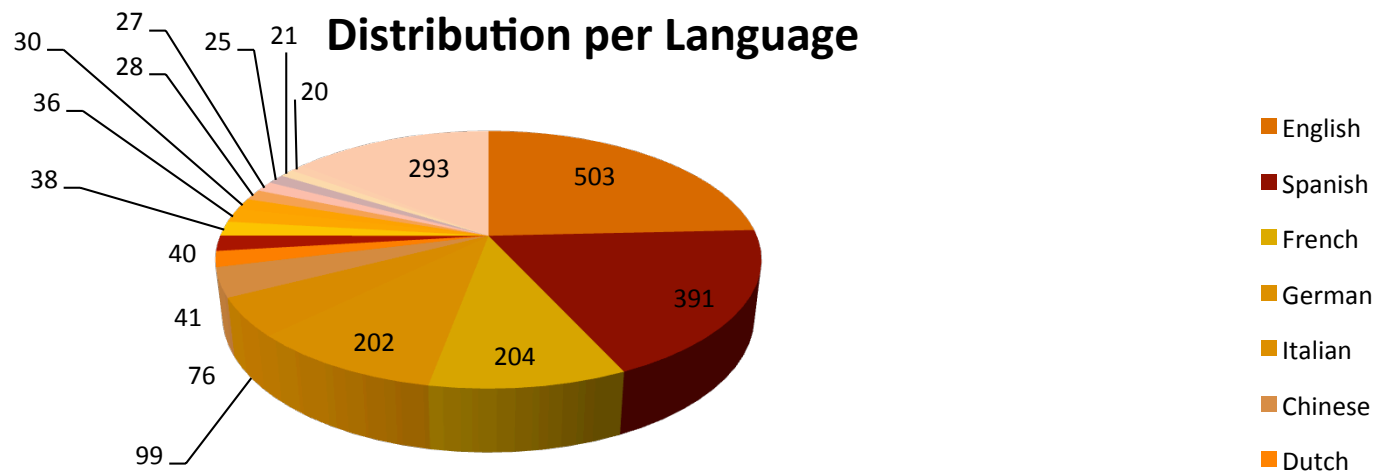
In the month(s) to come...

- Improved user and access rights management, catering for different roles in the resources lifecycle; production, maintenance, procurement

- Search engine optimisations

- Improved single resource view

- Recommendation services

- Fix usability bugs

- Data migration tools

In the month(s) to come...

- More META-SHARE nodes and respective language resources will be integrated

- Integration of ELRA supported initiatives, LRE Map, Language Library

- Adoption of the META-SHARE platform and framework by ELRA

- Mappings to other resource inventories, enhancing the information provision dimension

- Full deployment of the services of ELRA and members within the META-SHARE network – from software availability and maintenance to language resources storage and preservation as well as legal support

Q/A

Thank you very much!

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance