

The European Language Resources Association and The New paradigm/trend of data sharing

Khalid Choukri

ELRA/ELDA

55 Rue Brillat-Savarin, F-75013 Paris, France

Tel. +33 1 43 13 33 33 -- Fax. +33 1 43 13 33 30

choukri@elda.org

<http://www.elda.org/> or <http://www.elra.info/>

Quick ... overview

- (Quick reminder) The ELRA foundation and Mission and structure
 - Identification, Cataloguing, Description
 - Legal Support, Right negotiation, Licensing, and Distribution of LRs
- What can/will ELRA do for META-SHARE
- What can/should META-SHARE do for ELRA and LRs Community
- Conclusions and future evolution(s) plans

ELRA's Foundation & Mission

An Improved infrastructure for Data sharing & HLT evaluation

- Created in February 1995, Support & initial funding from the European Commission
- **The non for profit Association of users of Language Resources (LRs)**
 - Data sets, NLP Tools
 - 60 to 120 members (R&D and Industry, all over the world)
- **A Repository Center taking care of all issues related to LR from Discovery, clearing IPR to sharing and distribution, for R&D and Commercial purposes**
- Discovery and Identification of LR (LR repositories: ELRA Catalogue, Universal Catalogue, LRE Map, etc.)
- **The biggest event dealing with LR & Evaluation since 1998 (LRECs)**
- Operational body & over 15 permanent employees: ELDA
- Sustainability & Self-supported

Dimensions of the Human Languages

Status of ELRA catalogues

- Speech (Audio)
 - Audio/Video
 - Text inc. documents
 - Signs
 - Hand writing
 - Gestures ... pointing
 - Images, inc. OCR documents
 - Biometrics
 - Multimodal & Multimedia

 - **LRs & Evaluation Packages**
 - **Multilinguality**
- **ELRA Catalogue (>1200 LRs):**
<http://catalogue.elra.info/>
 - **Universal Catalogue: (>2000 LRs)**
<http://universal.elra.info/>
 - **LRE-MAP as of mid-2012**
over 4000+ identified items

 - **Distribution :**
 - balanced between R&D and Industry

Dimensions of the Human Languages

Status of ELRA catalogues – Legal aspects

- Legal issues :
 - IPR issues: ownership, copyright holder, authorship
 - Privacy, Ethics, ...
 - Licensing (In-House, Permissive e.g. CC, more restricted, think of business models)
 - Institutional users: academic organizations and commercial organizations
 - Usage: R&D, development of technology, (re)distribution, distribution of derived products, ...
- Need for permanent staff to acquire/capitalize on knowledge in Legal matters and in conducting such negotiations

What can/will ELRA do for META-SHARE

ELRA brings to META-SHARE

- Language Resources:
 - More than 60% of current meta-share resources (>1000 items)
 - More than 20 free resources every quarter (see second batch announced today)
 - More resources from the ELRA Universal Catalogue, LRE-MAP, ...
 - Comprehensive initial description framework (META-DATA)
 - Established Quality Control and Quality assessment methodologies and procedures
 - Deep knowledge of the field of Language Resources and all related issues

- Legal expertise on IPR and Licensing
 - Expertise on all aspects of IPR
 - Expertise on licensing LRs
 - Comprehensive set of licenses, used as background for META-SHARE Commons
 - An existing and strengthen Legal helpdesk (no more restricted to ELRA members)

ELRA brings to META-SHARE

- Contribution to the Software development of the meta-share engine
- Technical support/maintenance/evolution/upgrade ...
 - of the software package
 - Of the meta-data descriptions (may be push it to an ISO standard)

What can/will META-SHARE do for ELRA

What can/will META-SHARE

do for

~~ELRA~~

ELRA members & LR Community

What does META-SHARE brings to ELRA ?

A very efficient and sophisticated instrument to better render its services, Services as defined by the LR community in ELRA's core MISSION: Making LR usable under clean, clear legal terms, and fair conditions

What does META-SHARE brings to ELRA ?

- Provide *easy, uniform, one-step access to LRs*
 - through the aggregation of LR sources into one (common) catalogue
 - Facilitate the *LRs search and retrieval* processes
- Adoption of a distributed network of Repositories ... consolidated Partners
- Adoption of a *uniform* metadata schema
- Adoption of a common and trustable legal framework
 - Including a "standardized" set of licenses, reflecting ELRA expectations
- Services to assist resource owners in setting up (local/non-local) repositories
 - Both for internal use and networked ones
- ☑ **ELRA Catalogue to be Powered by META-SHARE engine**

- Stable, open sources, freely available package:
 - to set META-SHARE(-like) repositories
 - to incorporate more features (documentation/best practices storage)
 - Possibility to interoperate with collaborative work platforms
- Standardized version of Meta-data,
- Interfacing with the international initiative on ISLRN
 - International Standard for Language Resources Number
- More sophisticated Back-Office features

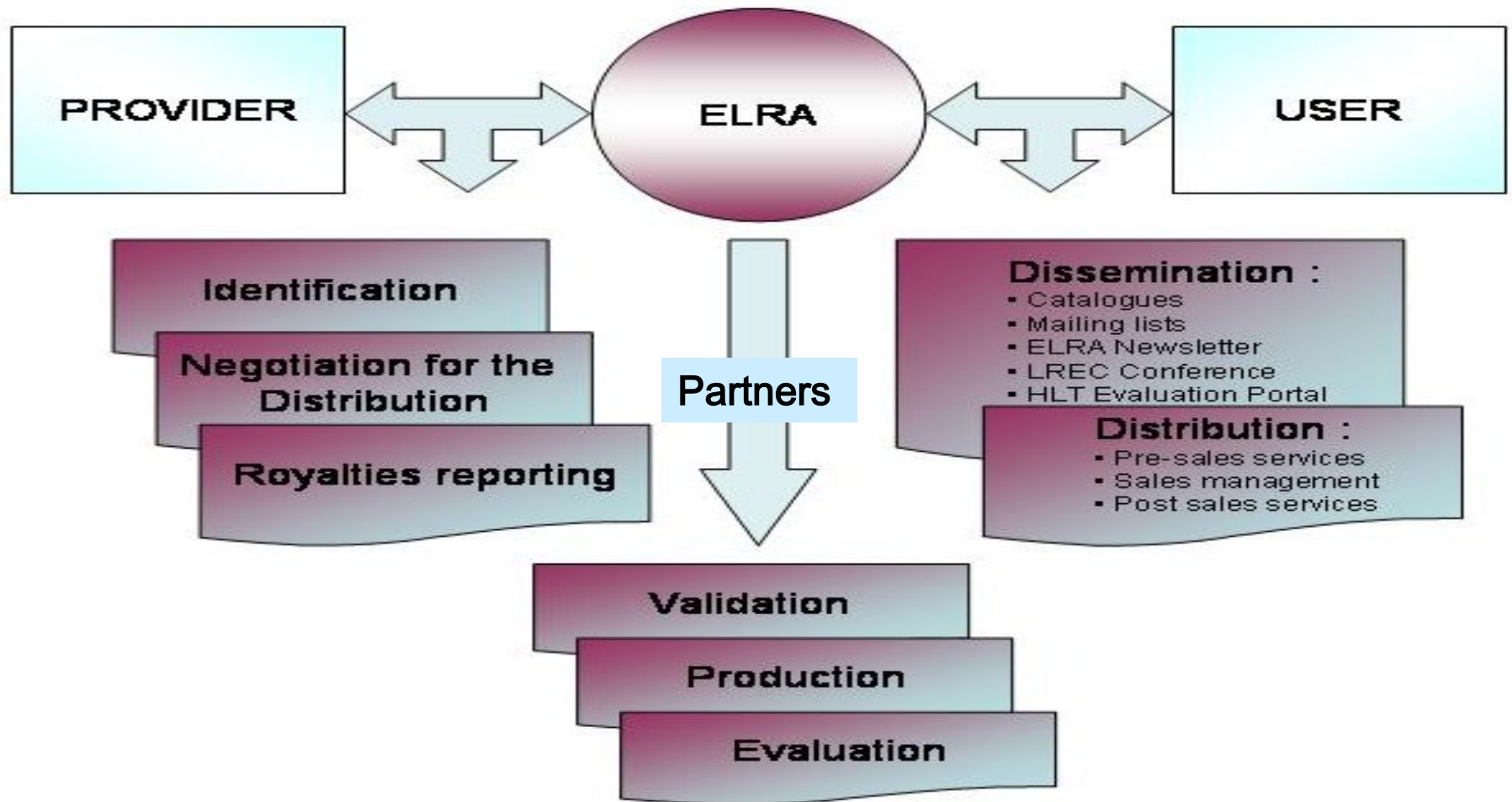
- Keep in place
 - Helpdesks (metadata, legal, technical)



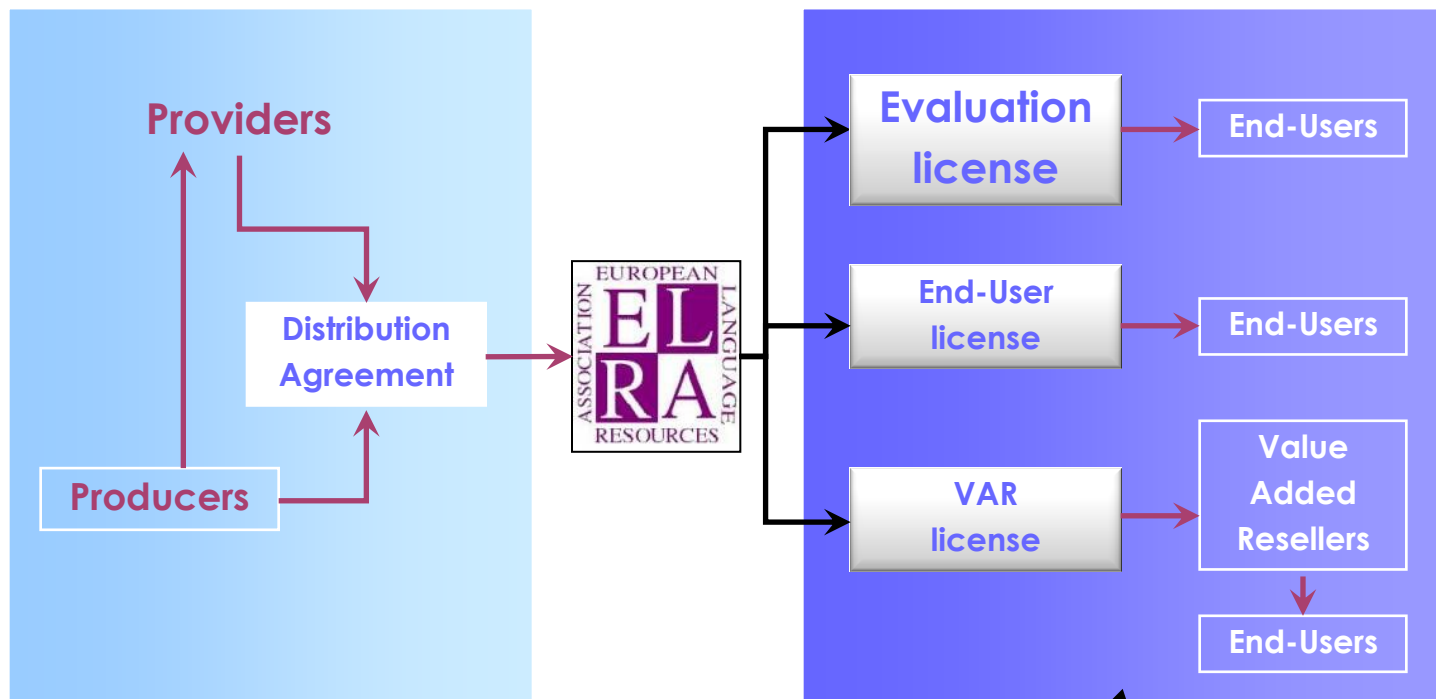
WISHES LONG LIFE TO META-SHARE

META SHARE

ELRA Activities... Identification, Distribution, Production of LR and HLT Evaluation & Dissemination



Example of Legal Issues - Licensing



From ELRA Licenses to the META-SHARE new set of licenses, even with larger scopes