

Legal Issues for Sharing Language Resources: Constraints and Best Practices



LREC Satellite workshop

17 May 2010 – Malta
@LREC 2010

Organising Committee

Khalid Choukri & Valérie Mapelli (ELRA / ELDA, Paris, France)

Denise DiPersio (Linguistic Data Consortium, USA)

Marc Kupietz (Institut für Deutsche Sprache, Germany)

With the support of:





WORKSHOP INTRODUCTION

*LRs & Legal issues,
a quick vista from 15 years of practices*

by

Khalid CHOUKRI

ELRA/ELDA

choukri@elda.org

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

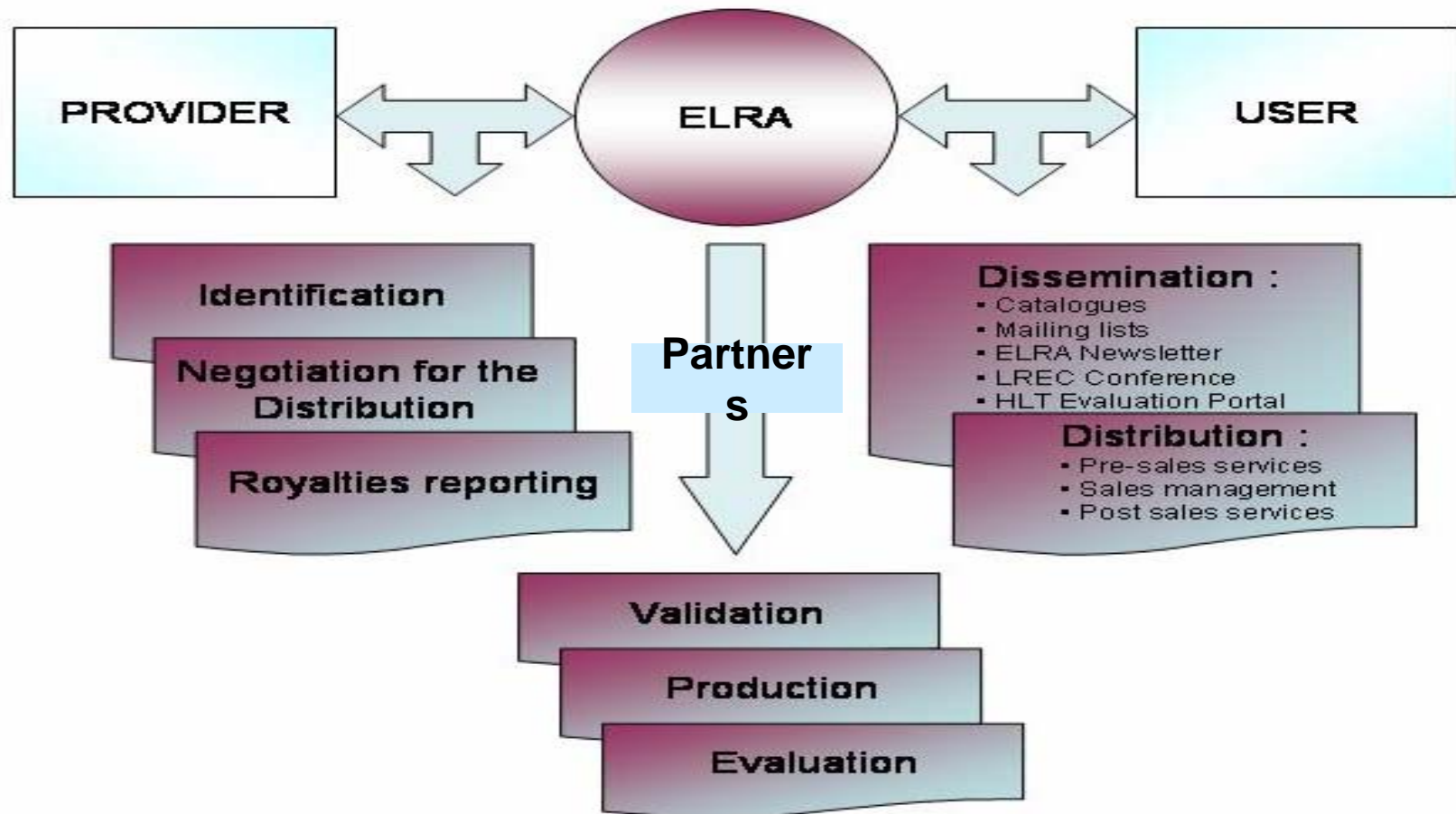
What are the “objects” under analysis



- **Language Resources : Data sets**
 - More linguistic than Content
- **Language processing tools & technologies & applications**
 - Taggers, tokenizers, signal analysers, etc.
- **Evaluation packages**
 - Datasets for testing, metrics for performance assessments, methodologies, comparable results, documentations, systems’ outputs, ...
- etc....

What are the “levels” involved?

Highlights through ELRA activities



What are the “levels” involved? Highlights through ELRA activities



- **Main Activities and legal repercussions/effects**
 - Identification (Collect and Dissemination of Information)
 - Distribution (Licensing) of (others) LRs but also sharing ours
 - Production of LRs from scratch ... from (our/others) background and/or know-how
 - Evaluation of Human Language Technologies, releasing rankings but also evaluation packages

Various aspects to scrutinize for LRs



- Types of rights ... Despite Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886, several rights are still in force:
 - Copyright (..... intellectual property)
 - Trademark
 - Patent
 - European Database Directive 96/9/EC
 - Ethical rules/ Privacy (e.g. CNIL in France → Privacy)
 - and more
 - Moral rights (e.g. under the French law)
 - Research/Education fair act(s)
- (often) Very national Even within the European Union

- What are the concerns & expectations from the HLT Community
- Demands, interests and desiderata from the linguistic versus content point of view
- IPR, data protection and ethical issues around Language Resources
- Technical measures for dealing with legal restrictions
- International diversity
- A licence that fits all or towards a uniform license model (existing differences and merging possibilities)
- New ideas and cooperation orientations

- Legal Rights to access to LRs (inc. Production of new data)
- (legal/licensing) Condition of use .. R&D vs Technology development (derived products)
- Re-distribution of the LRs (and derivatives)
- Pricing (free, for a fee, lump sum vs annual fee, etc.)
- ...etc

Rationales behind the Workshop



- Variety of legal issues: IPR (e.g. Copyrights), data protection issues, licensing schemas, etc.
- Variety of scenarios all along a Language Resource life cycle: collection, sharing with project partners, (re-)distribution.
- Obstacles:
 - lack of legal concern but also legal knowledge
 - multiplicity of IPR, data protection and liability issues
 - high number of home-made consent and license models
 - diversity of legal protection modes all over the world

- Sharing under the European regulation
 - Such resources can be shared only with players that apply a similar level of protection (ethic & privacy rules)
 - (or at least have similar laws regarding privacy & ethics)
 - According to the list drawn by the EC the USA does not comply with such rules.
 - » (Safe Harbor principles)

14:30 - 15:00 Copyrights, contracts, licenses, other legal issues
Isabelle Gavanon, FIDAL / ELDA consultant

15:00 - 15:30 Copyright and machine translation under United States law
Erik Ketzan, Attorney at Law

15:30 - 16:00 Creative Commons from different European national laws (UK, Greece, Norway, etc.)
Prodromos Tsiavos, Avgerinos & Partners

16:00 -16:30 Coffee break

16:30 - 17:00 Licensing and Sharing Language Resources: An Approach Inspired by Creative Commons and Open Science Data Movements

John Hendrik Weitzmann, Georg Rehm, Hans Uszkoreit,

Speaker: John Hendrik Weitzmann, EEAR - European Academy of Law and Computing

17:00 - 17h15 Some Implications of US Initiatives for 'Fair Research' and Open Access on the Development and Distribution of Language Resources

Denise DiPersio, LDC

17:15 - 17:30 Language resources and Amazon Mechanical Turk: legal, ethical and other issues

Gilles Adda, LIMSI

17:30 - 17:45 Processing Language Resources on the basis of the Consent

Christina Bankhardt, IDS Germany

17:45 - 18:30 LRs & Legal issues:

Panelists:

Isabelle Gavanon,

Erik Ketzan,

Prodromos Tsiavos,

John Hendrik Weitzmann

moderators : Khalid Choukri, ELRA/ELDA & Marc Kupietz, IDS Germany

➤ Let us start