Project no. **296371**

SAVAS

'Sharing AudioVisual language resources for Automatic Subtitling'

Collaborative Project
Information and Communication Technologies

Deliverable: D3.3.2

Legal and Licensing agreements: Update

Due date of deliverable: 31 10 2013
Actual submission date: 04 11 2013

Start date of project: 01/05/2012 Duration: 24 Months

Contact person responsible for this deliverable: Carlo Aliprandi
Organisation name responsible for this deliverable: Synthema

Revision [1.0]

Project co-funded by European Commission within the Seventh Framework Programme		
Dissemination level		
PU	Public	
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	X

DOCUMENT INFO

Author

Author	Company	E-mail
Carlo Aliprandi	Synthema	carlo.aliprandi@synthema.it
Arantza del Pozo	Vicomtech	adelpozo@vicomtech.org

Document Control

Document version #	Date	Change
V0.1	20 09 2013	Preliminary draft version
V0.2	18 10 2013	Added section on IPR questionnaire. Added METASHARE licenses
V0.3	24 10 2013	Added results of data collection from partners.
V0.4	25 10 2013	Updated licenses for resources
V0.5	29 10 2013	Added conclusions and executive summary
V1.0	30 10 2013	Revised version

Document Data

Point of Contact	Name: Carlo Aliprandi Partner: Synthema Address: Via Malasoma 24 - Pisa Phone: +39 050 9656411 Fax: +39 050 9656421 E-mail: carlo.aliprandi@synthema.it

TABLE OF CONTENTS

1. Executive Summary

As an update of deliverable D3.3.1, this document describes the progress made regarding the legal and licensing agreements linked to the sharing of the SAVAS results.

Section 2 details the licensing foundations settled down for data sharing. In Section 3, the questionnaire arranged to collect legal data and identify the licensing schema for each sharable SAVAS language resource is described. Finally, Section 4 presents the agreements reached.

2. Licensing foundations for data sharing

In the framework of the project, a legal expert was appointed with the specific goal of giving support in the definition and negotiation of IPR rights and licensing schemes to the SAVAS partners and data owners.

After investigating the wide range of existing licensing model schemes for data sharing, the legal expert recommended the adoption of the META-SHARE and Creative Commons (CC) schemes shown in Figures 1 and 2.

The first is a set ready to use commercial licenses pre-defined within the framework of META-SHARE by a team of legal experts to foster the needs of the Language Technologies community; the licenses are perpetual, worldwide, easily adoptable and well documented and are recommended by META-SHARE when licensing commercially to non META-SHARE members. By adopting these pre-set templates the owner (or in our case the SAVAS consortium) is spared the trouble of having to write a commercial agreement that fits his/her needs; thanks to the overview Scheme[1], simple questions help identify the desired combination and the text of the agreement can be downloaded and filled.

The second set, also recommended by META-SHARE, represents the best known and best documented scheme for open licensing, when sharing knowledge and creativity for non-commercial purposes. In this case too, accessible documentation and pre-defined templates make licensing to third parties straightforward.

The next step was to identify the most suitable set of licenses among them which could satisfy the SAVAS interests. In order to choose the most adequate licensing schemes, face to face meetings and discussions were held with each partner. In the end, the following three licenses were identified:

- the **META-SHARE Commercial NoReDistribution For-a-Fee** license[2] (C-NoReD-FF), for *commercial* purposes

- the **META-SHARE NonCommercial NoReDistribution For-a-Fee** license[3] (NC-NoReD-FF), for *R&D* purposes

---

[1] http://www.meta-net.eu/meta-share/meta-share-licenses/Tabular%20Presentation%20of%20No%20redistribution%20Licences.pdf

[2] http://www.meta-net.eu/meta-share/meta-share-licenses/META-SHARE%20Commercial%20NoRedistribution%20For-a-Fee_v0%207.pdf

- and the **Creative Commons Attribution NonCommercial NoDerivs** license[4] (CC BY NC SA), also for _R&D_ purposes

In a nutshell, language resources with the C-NoReD-FF licensing schema cannot be redistributed, require the payment of a fee and allow derivatives which can be used for commercial purposes. The NC-NoReD-FF license does not allow redistribution, also requires the payment of a fee and its derivatives can only be used for non-commercial purposes. Finally, resources with CC BY NC SA schemes cannot be redistributed either, require attribution, are free and allow derivatives which can only be used for non-commercial purposes and need to be shared under the same terms.

Licence name	Licence Attributes	Should the Original be redistributed within the META SHARE Network only?		Is Attribution required?		What purposes can the Original Resource be used for?		Are Derivatives allowed?		Do derivateives have to be shared under the same terms?		Is there any Fee for this Licence?	
		Yes	No	Yes	No	Commercial	NonCommercial	Yes	No	Yes	No	Yes	No
CC-ZERO			✓		✓	✓		✓			✓		✓
CC BY	BY		✓	✓		✓		✓			✓		✓
CC BY SA	BY SA		✓	✓		✓		✓		✓			✓
CC BY ND	BY ND		✓	✓		✓			✓	N/A	N/A		✓
CC BY NC SA	BY NC SA		✓	✓			✓	✓		✓			✓
CC BY NC	BY NC		✓	✓			✓	✓			✓		✓
CC BY NC ND	BY NC ND		✓	✓			✓		✓	N/A	N/A		✓

Figure 1. Creative Commons licensing schemes

---

[3] http://www.meta-net.eu/meta-share/meta-share-licenses/META-SHARE%20NonCommercial%20NoRedistribution%20For-a-Fee-v%201.0.pdf
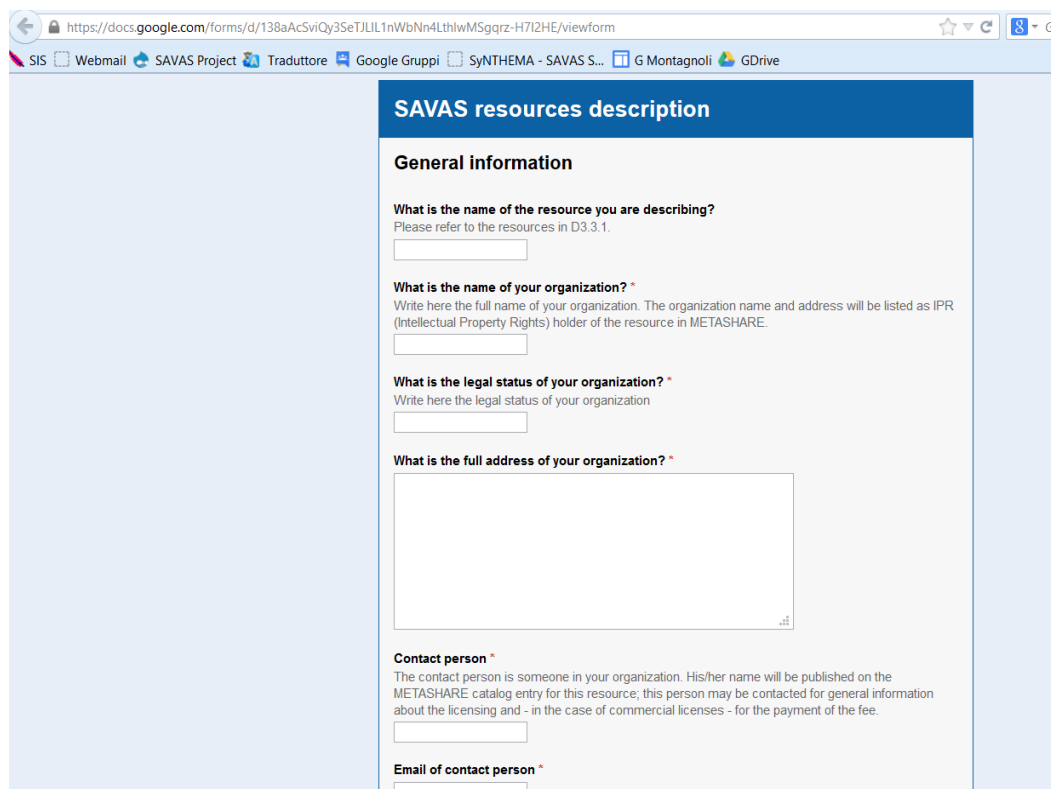
[4] http://creativecommons.org/licenses/by-nc-sa/3.0/

Licence name	Licence Attributes	Can the Original Resource be redistributed		What purposes can the Original Resource be used for?		Are Derivatives allowed?		Can Derivatives be commercially used?		Is there any Fee for this Licence?	
		Yes	No	Commercial	Non Commercial	Yes	No	Yes	No	Yes	No
META-SHARE Commercial NoRedistribution For-a-Fee	C-NoReD-FF		✓	✓		✓		✓		✓	
META-SHARE Commercial NoRedistribution	C-NoReD		✓	✓		✓		✓			✓
META-SHARE Commercial NoRedistribution NoDerivatives For-a-fee	C-NoReD-ND-FF		✓	✓			✓	N/A	N/A	✓	
META-SHARE Commercial NoRedistribution NoDerivatives	C-NoReD-ND		✓	✓			✓	N/A	N/A		✓
META-SHARE NonCommercial NoRedistribution For-a-Fee	NC-NoReD-FF		✓		✓	✓			✓	✓	
META-SHARE NonCommercial NoRedistribution	NC-NoReD		✓		✓	✓			✓		✓
META-SHARE NonCommercial NoRedistribution NoDerivatives For-a-fee	NC-NoReD-ND-FF		✓		✓		✓	N/A	N/A	✓	
META-SHARE NonCommercial NoRedistribution NoDerivatives	NC-NoReD-ND		✓		✓		✓	N/A	N/A		✓

Figure 2. META-SHARE licensing schemes

3. Legal and licensing questionnaire

In order to clean the legal status and identify the licensing scheme per sharable language resource, a questionnaire was designed and made available online as a Google form[5] to be filled by the SAVAS data providers for each of their owned data sets with the support of the legal expert. Figure 3 shows the initial page of the online questionnaire. The full set of questions can be found in Annex I. Answers to the questions were then used to identify the conditions and purposes under which the resources should be made available.

---

[5] https://docs.google.com/forms/d/138aAcSviQy3SeTJLIL1nWbNn4LthIwMSgqrz-H7I2HE/viewform

Figure 1. Legal and licensing questionnaire, online form

4. Data sharing agreements

Tables 1 and 2 list the legal and licensing agreements reached for each sharable SAVAS resource, per language.

As it can be seen in the table, the collection of legal data and the identification of the most adequate licenses for the resources owned by the data providers in the consortium has been achieved. The legal status has been cleared for almost all results owned by the SAVAS partners. Regarding licensing, a general consensus has been found on the following principles:

- Resources will be made available to any interested party (also to non META-SHARE members)
- Resource Attribution will be required
- Derivatives will be allowed
- "Share alike" licenses will be used for derivatives
- Commercial and Research licenses will be provided
- A fee will be charged for Commercial licenses
- Research licenses will be provided for free/for-a-fee

The only remaining discussion still ongoing within the consortium is linked to the level of restriction of the transcribed audio datasets for research purposes. While some partners are willing to share them for free and allow derivative works, others are concerned this may have a negative impact in the long term at exploitation level.

The consortium is aware of several potential competitors which are backed up by research centers who could easily get hold of the transcribed audio datasets and use them to improve the commercial systems of their partners, bringing to the market similar transcription and dictation systems as those developed within the project. More restrictive license schemes are being considered to avoid this, while allowing their use as evaluation benchmarks

The legal and licensing status of the resources owned by third-parties is still under discussion. Negotiations have been launched for those resources owned by third-parties (such as French audio or German crawled texts) and the corresponding licensing schemes are under discussion. We plan to advance these discussions during the next few months and expect to close legal and licensing negotiations for all the SAVAS resources before their distribution through the SAVAS META-SHARE repository, forecasted to be made available by the end of April 2014.

Basque		Language Resource		Owner	Amount	Commercial license	Research license
	Audio	Broadcast news		ETB	200h	C-NoReD-FF	CC BY NC SA
		Sports		ETB	20h	C-NoReD-FF	CC BY NC SA
	Text	Autocues, scripts and subtitles		ETB	125M	C-NoReD-FF	CC BY NC SA
		Crawled news		ETB	12M	C-NoReD-FF	CC BY NC SA
				THIRD PARTIES	341M	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III		MIXER-VICOMTECH	10h+5h+5h	C-NoReD-FF	Under discussion
Spanish	Audio	Broadcast news		ETB	200 h	C-NoReD-FF	CC BY NC SA
	Text	Autocues, scripts and subtitles		ETB	188M	C-NoReD-FF	CC BY NC SA
		Crawled news		ETB	17M	C-NoReD-FF	CC BY NC SA
				THIRD PARTIES	983M	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III		MIXER-VICOMTECH	10h+5h+5h	C-NoReD-FF	Under discussion
Portuguese	Audio	Interview/debate		RTP	19h	C-NoReD-FF	CC BY NC SA
	Transcribed audio	Test sets I, II and III		RTP-VOICEINTERACTION	1h+0,5h+0,5h	C-NoReD-FF	Under discussion
Italian	Audio	Broadcast news		RAI	150h	C-NoReD-FF	C-NoReD-FF
	Text	Autocues, scripts and subtitles		RAI	100M	C-NoReD-FF	C-NoReD-FF
		Crawled news		RAI	10M	C-NoReD-FF	C-NoReD-FF
				THIRD PARTIES	990M	Under discussion	Under discussion
		Crawled sports		THIRD PARTIES	500k	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III		RAI-SYNTHEMA	8h+4h+4h	C-NoReD-FF	Under discussion

Table 1. SAVAS data sharing agreements (I)

Swiss Italian	Audio	Broadcast news	SWISSTXT	50h	C-NoReD-FF	CC BY NC SA
	Text	Autocues, scripts and subtitles	SWISSTXT	5M	C-NoReD-FF	CC BY NC SA
		Crawled news	THIRD PARTIES	100M	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III	SWISSTXT-SYNTHEMA	2h+1h+1h	C-NoReD-FF	Under discussion
French	Audio	Broadcast news	THIRD PARTIES	150h	Under discussion	Under discussion
	Text	Crawled news	THIRD PARTIES	1B	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III	SYNTHEMA	8h+4h+4h	Under discussion	Under discussion
Swiss French	Audio	Broadcast news	SWISSTXT	50h	C-NoReD-FF	CC BY NC SA
	Text	Autocues, scripts and subtitles	SWISSTXT	4M	C-NoReD-FF	CC BY NC SA
		Crawled news	THIRD PARTIES	100M	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III	SWISSTXT-SYNTHEMA	2h+1h+1h	C-NoReD-FF	Under discussion
German	Audio	Broadcast news	THIRD PARTIES	150h	Under discussion	Under discussion
	Text	Crawled news	THIRD PARTIES	1B	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III	SYNTHEMA	8h+4h+4h	Under discussion	Under discussion
Swiss German	Audio	Broadcast news	SWISSTXT	50h	C-NoReD-FF	CC BY NC SA
	Text	Autocues, scripts and subtitles	SWISSTXT	5M	C-NoReD-FF	CC BY NC SA
		Crawled news	THIRD PARTIES	100M	Under discussion	Under discussion
	Transcribed audio	Test sets I, II and III	SWISSTXT-SYNTHEMA	2h+1h+1h	C-NoReD-FF	Under discussion

Table 2. SAVAS data sharing agreements (II)

5. Conclusions

This deliverable has reported the progress made in the legal and licensing agreements related to the sharing of the SAVAS language resources.

Advancements have been achieved leveraging on face to face discussions and negotiations with each result owner within the consortium. The designed Legal and Licensing Questionnaire has resulted to be a useful tool to collect the specific legal information and identify the most adequate license scheme per sharable language resource.

While the legal status and license types of almost all language resources owned by the SAVAS consortium has been cleared and defined, discussions and negotiations are still ongoing to agree on: (a) the most adequate license scheme of transcribed audio datasets for research purposes and (b) the legal and licensing status of the resources owned by third-parties external to the consortium.

The pending discussions are expected to be closed before launching the SAVAS META-SHARE repository by the end of April 2014.

Annex I – Legal and licensing questionnaire

1.1 General information

What is the name of the resource you are describing? Please refer to the resources in D3.3.1.

What is the name of your organization? *Write here the full name of your organization. The organization name and address will be listed as IPR (Intellectual Property Rights) holder of the resource in METASHARE.

What is the legal status of your organization? *Write here the legal status of your organization

What is the full address of your organization? *

Contact person *The contact person is someone in your organization. His/her name will be published on the METASHARE catalog entry for this resource; this person may be contacted for general information about the licensing and - in the case of commercial licenses - for the payment of the fee.

Email of contact person *

Telephone number of contact person

What is the domain of activities of your organization? *Your organization's core business.

[ ]

Additional IPR holders - if any Add here the name and address of other organizations/people that have joint ownership of the resource. The additional persons/organizations that you listing here must have agreed upon joint ownership.

[ ]

1.2 Parts

Is the material in object made up of different parts? *DEFAULT ANSWER: "no". Select "yes", if the resource that we asked to describe is actually made up of different parts for which different Intellectual Property rights hold, of for which you want to give a different licesing scheme. You will be asked to describe the different parts in which to split the resource and a new questionnaire for each of them will be issued.

◯ Yes

◯ No

[ ]

1.3 Describe parts

Please provide a description of the different parts in which to split the resource. *A new questionnaire will be sent to you with the different parts to be described separately.

[ ]

1.4 IPR

Does your organization hold the complete intellectual property of the material in object? *DEFAULT answer "yes"; select "no" only if other IPR holders exist, other than the ones that have been listed in the first part of this questionnaire, e.g. people

or organizations outside of the SAVAS consortium and who might object to the distribution of the material.

   ○ Yes

   ○ No

1.5 IPR info

Please provide us with the names of the person(s) holdin Intellectual Property Rights over the resource as well as some additional information on the Intellectual property status of the resource. *A new questionnaire will be issued based on your input.

1.6 Redistribution

Do you want to allow some groups to download also the resource under conditions (licence) of your choice within the SAVAS/METASHARE node? *DEFAULT ANSWER: "yes". If "no" is selected, only the metadata will be accessible and no further request about licence will be asked.

1.7 Licensing information

Here you enter the information that will allow us to chose the appropriate license under which you can distribute you resource in the SAVAS/METASHARE node.

Which users should be able to dowload you resource? *Within METASHARE the description of your resource (metadata) is visible to everyone. The download (under whatever condition you may choose) can be limited to specific users profiles. Please specify. DEFAULT ANSWER: "Everyone"

   ○ Everyone (including non Metashare members)

   ○ All Metashare members (including SAVAS)

   ○ Members of the SAVAS consortium only

Is attribution required? *If someone uses your resource must give your organization explicit credit (but not in any way that suggests that they endorse you or your use of the work). DEFAULT ANSWER: "Yes"

○ Yes

○ No

What purposes can the original material be used for? *Insert one or more of the possible values. DEFAULT ANSWER: "Research, Commercial". If "Non commercial" is selected, the material could be used to produce open-source/freely available tools.

☐ Research

☐ Commercial

☐ Non commercial

Are derivatives allowed? *In allowing derivatives, users may alter, transform, or build upon this work.
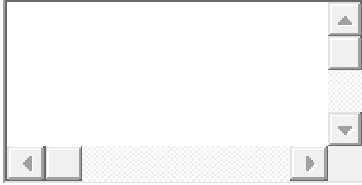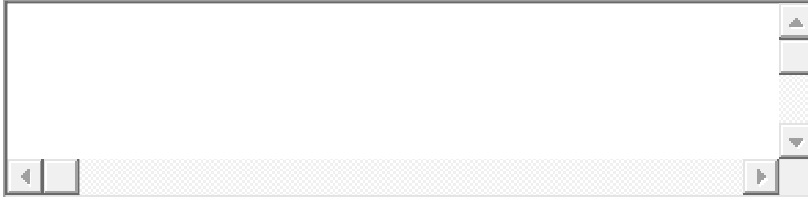
○ Yes

○ No

Can derivatives be commercially used? *

○ Yes

○ No

Do derivatives have to be shared under the same terms? *if you choose yes the "Share Alike" rule is apply: whoever alters, transforms, or builds upon this work, may distribute the resulting work only under the same or similar license to this one.

○ Yes

○ No

Should people pay a fee to download and use your resource? *If you answer yes, a commercial license will be chosen. You will be asked later to provide information about fees to be inserted in the SAVAS/METASHARE node. DEFAULT ANSWER: "Yes, only for commercial use".

○ Yes, always

Thanks a lot for your help. Your response has been recorded.