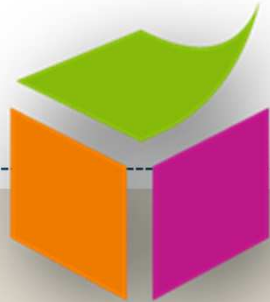


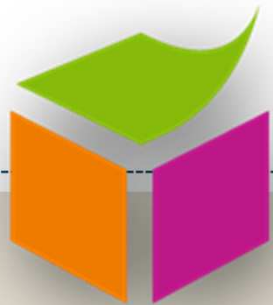
# Web sémantique et gestion des connaissances

## RDF – Mini projet **tri**plification



INTRODUCTION AU WEB DES DONNÉES (WEB SÉMANTIQUE)  
L'ARCHITECTURE DES STANDARDS DU W<sub>3</sub>C DU WEB  
SÉMANTIQUE (DU XML AU RDF)  
LES DIFFÉRENTES SYNTAXES DE SÉRIALISATION DE RDF

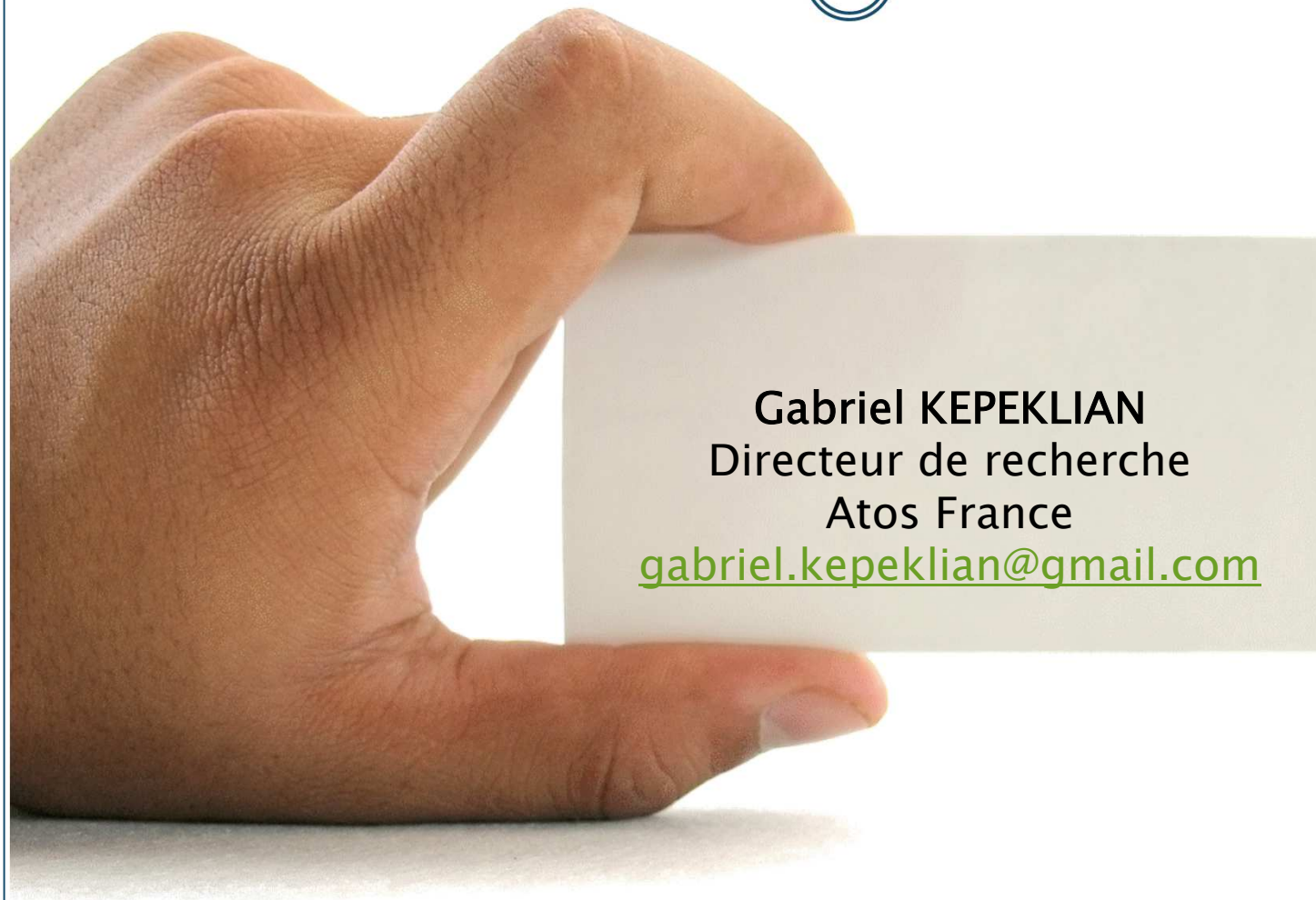
# Semantic web and Knowledge management



## RDF – **Triplification** short project

2

**INTRODUCTION TO THE WEB OF DATA (SEMANTIC WEB)  
THE ARCHITECTURE OF THE SEMANTIC WEB  
W<sub>3</sub>C STANDARDS (FROM XML TO RDF)  
THE DIFFERENT SERIALISATION SYNTAXES OF RDF**



# Short project

4



# Nommage et rendu

5

- J'attends un fichier ZIP
  - WS-NOM-Prénom.zip
- Pas de fichiers séparés
- Pour mardi 7 déc.
- envoi sur [gabriel.kepeklian@gmail.com](mailto:gabriel.kepeklian@gmail.com)  
en utilisant grosfichiers.com
- Pas de Mail

# Cahier des charges

6

## DÉVELOPPEMENT D'UN TRIPLIFICATEUR

# Specifications

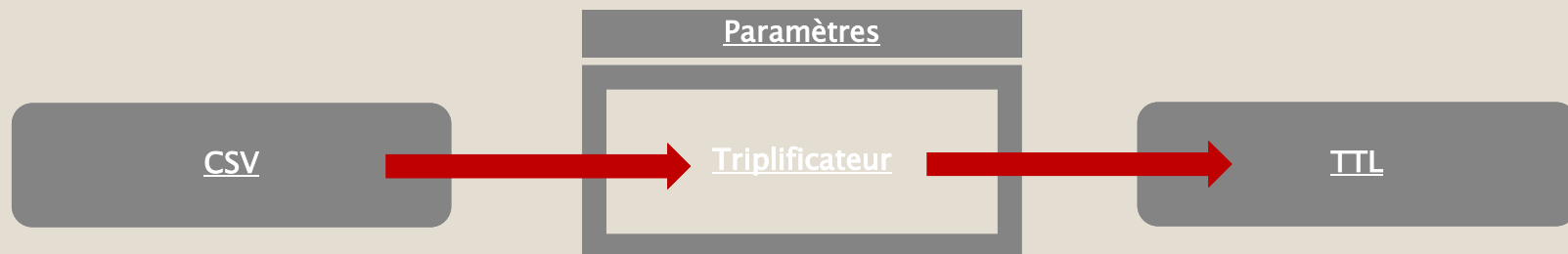
7

## DEVELOPMENT OF A TRIPLIFIER

# Cahier des charges

8

- Former des groupes de 3 à 4 étudiants (ni plus ni moins)
- Développer un outil qui prend en entrée un fichier CSV et le transforme en un fichier contenant les triplets RDF équivalents dans une sérialisation Turtle.

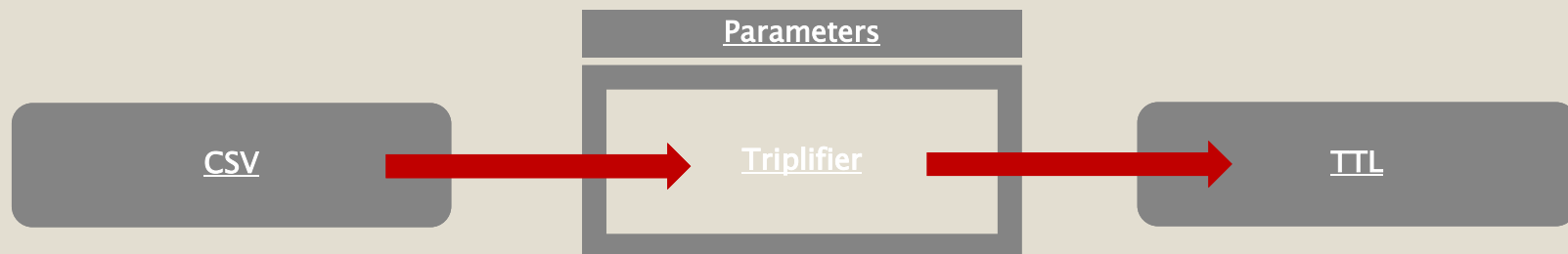




# Specifications

9

- Form groups of 3 to 4 students (no more and no less)
- Develop a tool that takes a CSV file as input and transforms it into a file containing the equivalent RDF triples in a Turtle serialization.



# Quatre jeux d'essai

10

- Votre code doit fonctionner pour les 4 jeux d'essai fournis :
  - test1.csv
  - test2.csv
  - test3.csv
  - test4.csv
- Regardez les bien pour identifier ce qui les caractérisent et les cas de figures que vous devez bien prendre en compte.

# Four test sets

11

- Your code must work for the 4 test sets provided:
  - test1.csv
  - test2.csv
  - test3.csv
  - test4.csv
- Take a good look at them to identify what characterizes them and what you need to take into account.

# Spécifications du fichier à triplifier

12

- Les fichiers CSV présentent des caractéristiques. Certaines peuvent-être choisies par l'utilisateur. Des options par défaut existent et sont proposées.
  - L'utilisateur choisit le fichier à traiter parmi les fichiers (.csv)
  - C'est un fichier unicode
- Ligne de titres
  - par défaut, votre application prend en compte "avec titres"
  - quand il y en a une ligne de titres, elle est unique, et c'est nécessairement la première ligne non vide
  - l'utilisateur peut spécifier le numéro de la ligne de titres (un nombre positif !)
  - il peut y avoir des lignes vides avant la ligne de titres
  - les données sont toujours placées après la ligne de titres quand il y en a une
- Lignes de données
  - il peut y avoir des lignes vides avant la première ligne de données
  - l'utilisateur peut spécifier le séparateur (parmi : espace, tab, virgule, point-vigule, barre verticale) – par défaut, l'application prend en compte ce qui est défini dans le fichier d'initialisation (.ini)
  - l'utilisateur peut spécifier le numéro de la première ligne de données
  - les données sont lues comme des chaînes de caractères unicode

# Specifications of the file to be triplicated

13

- CSV files have some characteristics. Some of them can be chosen by the user. Default options exist and are proposed.
  - The user chooses the file to process among the files (.csv)
  - It is a unicode file.
- Title line
  - by default, your application takes into account "with titles".
  - when there is a title line, it is unique, and it is necessarily the first non-empty line.
  - the user can specify the number of the title line (a positive number!).
  - there can be empty lines before the title line.
  - the data are always placed after the title line when there is one.
- Data lines
  - there can be empty lines before the first line of data.
  - the user can specify the separator (among: space, tab, comma, semi-colon, vertical bar) - by default, the application takes into account what is defined in the initialization file (.ini).
  - the user can specify the number of the first line of data.
  - the data are read as unicode strings.

# Spécifications du fichier créé

14

- En sortie :
  - C'est un fichier Turtle qui respect parfaitement la syntaxe Turtle.
  - Son nom peut être donné par l'utilisateur - Par défaut, c'est le même nom que le fichier en entrée mais avec l'extension .ttl
  - Il utilise deux préfixes. Un pour les sujets, un autre pour les prédicats. La définition des préfixes peut être donnée par l'utilisateur (dans l'IHM) – par défaut, ce sont des valeurs qui sont dans un fichier de paramètres (.ini) lu par l'application
  - L'application ne produit pas de lignes vides.
  - C'est un fichier unicode.

# Specifications of the created file

15

- The output:
  - This is a Turtle file that follows the Turtle syntax perfectly.
  - Its name can be given by the user - By default, it is the same name as the input file but with the extension .ttl
  - It uses two prefixes. One for subjects, another for predicates. The definition of the prefixes can be given by the user (in the GUI) - by default, they are values that are in a parameter file (.ini) read by the application
  - The application does not produce empty lines.
  - It is a unicode file.

# Initialisation du triplificateur

16

- Un fichier d'initialisation (.ini) permet de fixer les valeurs par défaut. Voici à quoi il se conforme.

```
# noms : Jean Dupond, Pierre Martin, etc.  
# date : 3 décembre 2021  
# commentaire  
title 1      # numéro par défaut de la ligne de titres  
data 2      # numéro par défaut de la première ligne de données  
sep ", "    # par défaut, le séparateur est la virgule  
prefix_data d: <http://ex.org/data/>  
prefix_predicat p: <http://ex.org/pred#>
```



# Initialization of the triplicator

17

- An initialization file (.ini) is used to set the default values. Here is what it conforms to.

```
# nom : Jean Dupond
# date : 3 décembre 2021
# commentaire
title 1      # numéro par défaut de la ligne de titres
data 2       # numéro par défaut de la première ligne de données
sep ", "     # par défaut, le séparateur est la virgule
prefix_data d: <http://ex.org/data/>
prefix_predicat p: <http://ex.org/pred#>
```

# IHM du triplificateur

18

- Choix du fichier à traiter par parcours dans les répertoires
- Le N° de la ligne de titres
  - la valeur par défaut préremplie le choix
- Le N° de la première ligne de données à traiter
  - par défaut la ligne qui suit celle des titres
- Le N° de la dernière ligne de données à traiter
  - Par défaut, l'application traite les lignes jusqu'à la fin du fichier
- Le séparateur par défaut est celui du fichier (.ini), mais l'utilisateur peut donner une autre valeur (parmi : espace, tab, virgule, point-vigule, barre verticale)
- Les deux préfixes par défaut sont ceux du fichier (.ini), mais l'utilisateur peut donner une autre valeur
- L'utilisateur peut ouvrir le fichier en entrée pour le voir
- L'utilisateur peut ouvrir le fichier en sortie pour le voir

# Triplicator Interface

19

- Choice of the file to be processed by browsing through the directories.
- The number of the title line.
  - the default value pre-fills the choice.
- The number of the first line of data to be processed.
  - by default the line which follows that of the titles.
- The N° of the last line of data to be treated.
  - By default, the application processes the lines until the end of the file.
- The default separator is the one of the file (.ini), but the user can give another value (among : space, tab, comma, semi-colon, vertical bar).
- The two default prefixes are those of the file (.ini), but the user can give another value.
- The user can open the input file to see it.
- The user can open the output file to see it.

# Développement et technologie

20

- Le code source doit être commenté.
  - Au début du code, votre cartouche donne a minima votre nom et la date.
  - Dans le code, les commentaires expliquent les étapes du traitement
  - Les variables importantes sont commentées
- Une interface utilisateur WEB (pas de mode "ligne de commandes").
- Les dialogues et messages d'erreur sont en anglais.
- Technologie : Python

# Development and technology

21

- The source code must be commented.
  - At the beginning of the code, your title block gives at least your name and the date.
  - In the code, the comments explain the processing steps.
  - The important variables are commented.
- A WEB user interface (no "command line" mode).
- Dialogs and error messages are in English.
- Technology : Python.

# Votre rendu

22

- **A - Le code source, le fichier d'initialisation**
  - Python, vous livrez le code source documenté
- **B - Le tutoriel d'installation (word)**
  - Il prend la forme d'un Howto. Chaque étape est décrite, vous ne supposez pas que votre utilisateur connaît la technologie utilisée. Les liens pour télécharger le framework, sa documentation, etc. Vous n'oubliez rien.
  - Par exemple, si vous utilisez Anaconda, vous expliquez comment installer et utiliser Anaconda pour exécuter votre application.
- **C - Le mode d'emploi de l'application (word avec visuels)**
  - Il prend la forme d'un Howto. Vous guidez l'utilisateur sur au moins un exemple complet. Vous expliquez l'utilisation des options.
- **D - Les 4 fichiers résultats obtenus à partir des 4 jeux d'essai fournis.**
- **L'ensemble des fichiers (de A à D) sont placés dans un unique ZIP.**
- **Facultatif : une recherche pour faire un état de l'art. Si vous avez trouvé des articles relatifs à la triplification, joignez-les à votre rendu.**

# Your return

23

- **A - The source code, the initialization file.**
  - Python, you deliver the documented source code.
- **B - The installation tutorial (word).**
  - It takes the form of a Howto. Each step is described, you do not assume that your user knows the technology used. The links to download the framework, its documentation, etc. You don't forget anything.
  - For example, if you use Anaconda, you explain how to install and use Anaconda to run your application.
- **C - The application manual (word with visuals).**
  - It takes the form of a Howto. You guide the user through at least one complete example. You explain the use of the options.
- **D - The 4 result files obtained from the 4 test sets provided.**
- All the files (from A to D) are placed in a single ZIP.
- **Optional:** a search for a state of the art. If you have found any articles related to triplication, attach them to your rendering.

# Thank you

24