



Improving Data Quality for Risk Assessment

49th Focal Point meeting

GP/EFSA/ENCO/2018/03 GA

Pedro Nabais

- For the past years we have been working with EFSA on report of food data for the domains of microbiological, chemical contaminants and food additives.
- Some progress was been made when it comes to automate most processes related to the SSD data transmission;
- However, several problems persisted which made it difficult to accomplish this task



This project emerged from two distinct problems regarding data report:

- Technicians/Data experts, whether they are working on field sampling data or cleaning and preparing data for report, an absurd amount of manual work is done: too much time is spent;
- The more manual work we have, the more prone to errors the data will become.



Auto de Colheita de Amostras

Produtos Alimentares

1. Dados relativos à colheita
Data: ____/____/____ Hora: ____:____ NÚT: ____

2. ID Funcionário: Nome: _____ N.º: _____ PL de origem: _____

3. Identificação do operador económico
Designação Social: _____ Registo/Aprovação n.º: _____
Endereço Sede Social: _____
CP: _____ NIPC: _____ Código (CAE): _____
Designação Estab.: _____
Endereço: _____
CP: _____ Tel.: _____ Fax: _____
Distrito: _____ Concelho: _____ e-mail: _____ @
Nome: _____ Função: _____
Endereço: _____ CP: _____
Filiação: _____ B/C: _____
3.1 Identificação do Estabelecimento
3.2 Identificação do Representante Legal
Data Nascimento: ____/____/____ Naturalidade: _____
Nacionalidade: _____ Concelho: _____
Estado civil: _____ Freguesia: _____

4. Colheita da amostra
4.1 Ambiente: ☐ F. PNC ☐ F. Outro _____
4.2 Amostra: ☐ F. Única ☐ F. Duplicado ☐ F. Triplicado Quant.: _____ Lote: _____
☐ F. Aleatoriamente ☐ F. Homogeneização ☐ F. Outro N.º exempl. colhidos: _____
M / HF: n.º _____ Data Durab. Min./Data Lim. Consumo: ____/____/____

5. Ponto de colheita
☐ F. Armazenista ☐ F. Restauração ☐ F. Produção primária ☐ F. Embalador
☐ F. Detalho ☐ F. Venda ambulante ☐ F. Distribuição e Transporte ☐ F. Catering
☐ F. Indústria ☐ F. Outro _____ Qual? _____
Tipo de atividade: _____

6. Descrição da amostra
Designação da amostra: _____
Respon. pela introdução no mercado: _____
Origem: _____ Marca: _____ Temperatura de conservação: _____ °C
Local de exposição: _____ Preço/Unid.: _____ € Rótulo de _____ F. Sim
Quantidade exposta: _____ Preço/Tot.: _____ € origem _____ F. Não

7. Método de produção
7.1. Produtos de Origem Animal: ☐ F. 7.1.1. Pescado/Aquicultura ☐ F. 7.1.2. Moluscos/Bivalves ☐ F. 7.1.3. Selvagem ☐ F. 7.1.4. Água doce ☐ F. 7.1.5. DOC ☐ F. 7.1.6. DOP ☐ F. 7.1.7. VQPRD ☐ F. 7.1.8. IGP ☐ F. 7.1.9. Proteção integrada ☐ F. 7.1.10. Tradicional ☐ F. 7.1.11. Desconhecido
Local captura: _____
7.2. Produtos de Origem Vegetal: ☐ F. 7.2.1. Hortícolas ☐ F. 7.2.2. Ar livre ☐ F. 7.2.3. Estufa ☐ F. 7.2.4. Outro ☐ F. 7.2.5. Desconhecido
Qual: _____

8. Tipo de processamento
☐ F. Sem processamento ☐ F. Fermentação ☐ F. Porcelana/Cerâmica ☐ F. Tetra-Pak ☐ F. Inox
☐ F. Desconhecido ☐ F. Esterilização ☐ F. Plástico/Filme Plástico ☐ F. Papel de Cera ☐ F. Aço
☐ F. Pasteurização/UHT ☐ F. Fumagem ☐ F. Atmosfera Modificada ☐ F. Folha-de-flandres ☐ F. Tecido
☐ F. Concentração ☐ F. Salga ☐ F. Filmes termomoldáveis (seal) ☐ F. Filme e poliestireno ☐ F. Granel
☐ F. Desidratação/Secagem ☐ F. Folha de alumínio ☐ F. Filme e papel ☐ F. Vidro
☐ F. Outro ☐ F. Papel/Cartão ☐ F. Filme e alumínio ☐ F. Vácuo
Qual: _____
☐ F. Bag in box ☐ F. Cimento ☐ F. PET
☐ F. Madeira ☐ F. Outro _____

9. Tipo de acondicionamento
☐ F. Sem processamento ☐ F. Fermentação ☐ F. Porcelana/Cerâmica ☐ F. Tetra-Pak ☐ F. Inox
☐ F. Desconhecido ☐ F. Esterilização ☐ F. Plástico/Filme Plástico ☐ F. Papel de Cera ☐ F. Aço
☐ F. Pasteurização/UHT ☐ F. Fumagem ☐ F. Atmosfera Modificada ☐ F. Folha-de-flandres ☐ F. Tecido
☐ F. Concentração ☐ F. Salga ☐ F. Filmes termomoldáveis (seal) ☐ F. Filme e poliestireno ☐ F. Granel
☐ F. Desidratação/Secagem ☐ F. Folha de alumínio ☐ F. Filme e papel ☐ F. Vidro
☐ F. Outro ☐ F. Papel/Cartão ☐ F. Filme e alumínio ☐ F. Vácuo
Qual: _____
☐ F. Bag in box ☐ F. Cimento ☐ F. PET
☐ F. Madeira ☐ F. Outro _____

10. Outras informações
Os produtos constitutivos da amostra foram colocados em _____, próprios para o efeito, invioláveis e abertos apenas no momento da colheita. Foi-lhe atribuído o código (do funcionário) n.º _____ e foi selada com o(s) selo(s) n.º(s) _____, e as respectivas etiquetas/bolhas de segurança rubricadas por mim, pela(s) testemunha(s) e pela pessoa presente atrás identificada.
☐ F. Não tendo ficado qualquer exemplar em poder desta _____
Tendo ficado um exemplar em poder desta, que declarou tê-la recebido, após o que foi advertida de que é responsável pela guarda do mesmo, não podendo dele dispor antes de lhe ser notificado o resultado. _____
O produto encontrava-se à temperatura de _____ °C tendo sido colocados em:
☐ F. Refrigeração _____ °C ☐ F. Conservado/Transportado ☐ F. mala térmica ☐ F. Com _____ termoacumuladores ☐ F. outra frigorífica

11. Prova
Documental: _____
Testemunhal: _____
Outra: _____
Observações: _____

12. Feito do Auto
Para constar, se lavrou o presente auto que foi por mim elaborado e integralmente revisto nos termos do artigo 94.º do Código de Processo Penal, e que foi lido e achado conforme vai ser devidamente assinado.
A pessoa identificada no ato, _____
A testemunha, _____
A testemunha, _____
O funcionário, _____

REGISTO DE COLHEITA DE AMOSTRAS - RCA 3

N.º de file. _____

Localização: _____

1. Dados de identificação
Data: ____/____/____ Hora de início: ____:____ Local: _____

2. Identificação da amostra
Designação: _____ Origem: _____ Marca: _____ Temperatura de conservação: _____ °C
Local de exposição: _____ Preço/Unid.: _____ € Rótulo de _____ F. Sim
Quantidade exposta: _____ Preço/Tot.: _____ € origem _____ F. Não

3. Identificação do operador económico
Designação Social: _____ Registo/Aprovação n.º: _____
Endereço Sede Social: _____
CP: _____ NIPC: _____ Código (CAE): _____
Designação Estab.: _____
Endereço: _____
CP: _____ Tel.: _____ Fax: _____
Distrito: _____ Concelho: _____ e-mail: _____ @
Nome: _____ Função: _____
Endereço: _____ CP: _____
Filiação: _____ B/C: _____
3.1 Identificação do Estabelecimento
3.2 Identificação do Representante Legal
Data Nascimento: ____/____/____ Naturalidade: _____
Nacionalidade: _____ Concelho: _____
Estado civil: _____ Freguesia: _____

4. Colheita da amostra
4.1 Ambiente: ☐ F. PNC ☐ F. Outro _____
4.2 Amostra: ☐ F. Única ☐ F. Duplicado ☐ F. Triplicado Quant.: _____ Lote: _____
☐ F. Aleatoriamente ☐ F. Homogeneização ☐ F. Outro N.º exempl. colhidos: _____
M / HF: n.º _____ Data Durab. Min./Data Lim. Consumo: ____/____/____

5. Ponto de colheita
☐ F. Armazenista ☐ F. Restauração ☐ F. Produção primária ☐ F. Embalador
☐ F. Detalho ☐ F. Venda ambulante ☐ F. Distribuição e Transporte ☐ F. Catering
☐ F. Indústria ☐ F. Outro _____ Qual? _____
Tipo de atividade: _____

6. Descrição da amostra
Designação da amostra: _____
Respon. pela introdução no mercado: _____
Origem: _____ Marca: _____ Temperatura de conservação: _____ °C
Local de exposição: _____ Preço/Unid.: _____ € Rótulo de _____ F. Sim
Quantidade exposta: _____ Preço/Tot.: _____ € origem _____ F. Não

7. Método de produção
7.1. Produtos de Origem Animal: ☐ F. 7.1.1. Pescado/Aquicultura ☐ F. 7.1.2. Moluscos/Bivalves ☐ F. 7.1.3. Selvagem ☐ F. 7.1.4. Água doce ☐ F. 7.1.5. DOC ☐ F. 7.1.6. DOP ☐ F. 7.1.7. VQPRD ☐ F. 7.1.8. IGP ☐ F. 7.1.9. Proteção integrada ☐ F. 7.1.10. Tradicional ☐ F. 7.1.11. Desconhecido
Local captura: _____
7.2. Produtos de Origem Vegetal: ☐ F. 7.2.1. Hortícolas ☐ F. 7.2.2. Ar livre ☐ F. 7.2.3. Estufa ☐ F. 7.2.4. Outro ☐ F. 7.2.5. Desconhecido
Qual: _____

8. Tipo de processamento
☐ F. Sem processamento ☐ F. Fermentação ☐ F. Porcelana/Cerâmica ☐ F. Tetra-Pak ☐ F. Inox
☐ F. Desconhecido ☐ F. Esterilização ☐ F. Plástico/Filme Plástico ☐ F. Papel de Cera ☐ F. Aço
☐ F. Pasteurização/UHT ☐ F. Fumagem ☐ F. Atmosfera Modificada ☐ F. Folha-de-flandres ☐ F. Tecido
☐ F. Concentração ☐ F. Salga ☐ F. Filmes termomoldáveis (seal) ☐ F. Filme e poliestireno ☐ F. Granel
☐ F. Desidratação/Secagem ☐ F. Folha de alumínio ☐ F. Filme e papel ☐ F. Vidro
☐ F. Outro ☐ F. Papel/Cartão ☐ F. Filme e alumínio ☐ F. Vácuo
Qual: _____
☐ F. Bag in box ☐ F. Cimento ☐ F. PET
☐ F. Madeira ☐ F. Outro _____

9. Tipo de acondicionamento
☐ F. Sem processamento ☐ F. Fermentação ☐ F. Porcelana/Cerâmica ☐ F. Tetra-Pak ☐ F. Inox
☐ F. Desconhecido ☐ F. Esterilização ☐ F. Plástico/Filme Plástico ☐ F. Papel de Cera ☐ F. Aço
☐ F. Pasteurização/UHT ☐ F. Fumagem ☐ F. Atmosfera Modificada ☐ F. Folha-de-flandres ☐ F. Tecido
☐ F. Concentração ☐ F. Salga ☐ F. Filmes termomoldáveis (seal) ☐ F. Filme e poliestireno ☐ F. Granel
☐ F. Desidratação/Secagem ☐ F. Folha de alumínio ☐ F. Filme e papel ☐ F. Vidro
☐ F. Outro ☐ F. Papel/Cartão ☐ F. Filme e alumínio ☐ F. Vácuo
Qual: _____
☐ F. Bag in box ☐ F. Cimento ☐ F. PET
☐ F. Madeira ☐ F. Outro _____

10. Outras informações
Os produtos constitutivos da amostra foram colocados em _____, próprios para o efeito, invioláveis e abertos apenas no momento da colheita. Foi-lhe atribuído o código (do funcionário) n.º _____ e foi selada com o(s) selo(s) n.º(s) _____, e as respectivas etiquetas/bolhas de segurança rubricadas por mim, pela(s) testemunha(s) e pela pessoa presente atrás identificada.
☐ F. Não tendo ficado qualquer exemplar em poder desta _____
Tendo ficado um exemplar em poder desta, que declarou tê-la recebido, após o que foi advertida de que é responsável pela guarda do mesmo, não podendo dele dispor antes de lhe ser notificado o resultado. _____
O produto encontrava-se à temperatura de _____ °C tendo sido colocados em:
☐ F. Refrigeração _____ °C ☐ F. Conservado/Transportado ☐ F. mala térmica ☐ F. Com _____ termoacumuladores ☐ F. outra frigorífica

11. Prova
Documental: _____
Testemunhal: _____
Outra: _____
Observações: _____

12. Feito do Auto
Para constar, se lavrou o presente auto que foi por mim elaborado e integralmente revisto nos termos do artigo 94.º do Código de Processo Penal, e que foi lido e achado conforme vai ser devidamente assinado.
A pessoa identificada no ato, _____
A testemunha, _____
A testemunha, _____
O funcionário, _____

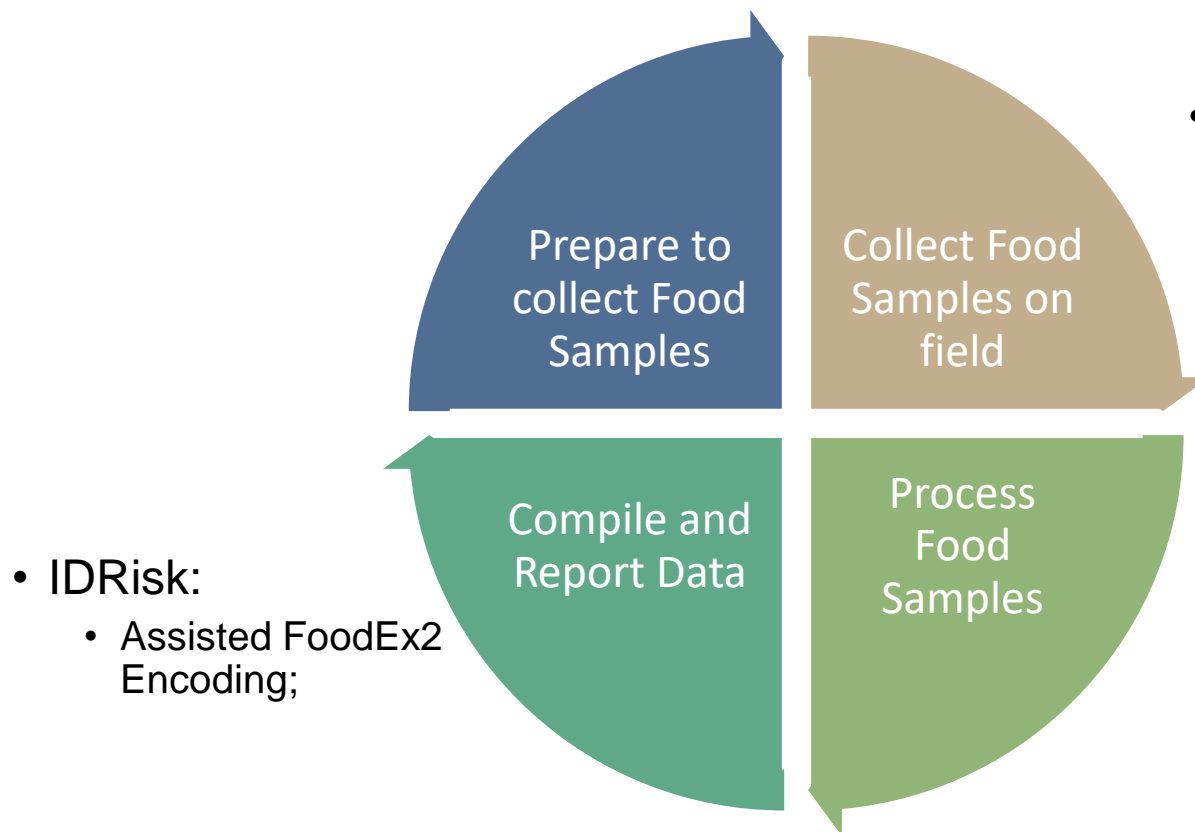
- A consortium between 3 partners (ASAE, INSA & HAPIH) was created
- They share the common interest of further developing their own official control National Data Management Systems (NDMS)
- Both ASAE and HAPIH are interested in implementing real-time sample data collection based on preparatory digital forms
- They are committed to investigate and implement an automatic approach to FoodEx2 classification of food samples using the knowledge and the existing databases



Main objectives

- Improve/Restructure the dynamic sampling forms module;
- Develop the application that will run on the mobile devices;
- Plan and implement an automatic NDMS FoodEx2 classification system for sampling descriptions.

Improve Data Quality!



- IDRisk:
 - Assisted FoodEx2 Encoding;

- IDRisk:
 - Digital Forms;
 - Data collect by mobile devices
 - Automated data insertion into systems;
 - Acquisition of other types of information;



Digital Form Application: Development

- The system is suitable for the creation of electronic forms, their management and their use;
- The creation of the forms are done in a WYSIWYG principle, i.e., the final format would be the same as the printed form:

Vinyl / Paper Cutting Order Form 

Name: <input type="text" value="test_4"/>	Submitted Date: <input type="text" value="000_1"/>	Needed by Date: <input type="text" value="000_1"/>
Email: <input type="text" value="email_0"/>		
Phone: <input type="text" value="phone_0"/>		

File Info

Folder File Name(s):

Illustration File:

Output Dimensions (L" x W"):

All Type outline?

☐ Paper Cutting

☐ Vinyl Cutting
 ☐ Weeding
 ☐ Taping
 ☐ Weed and tape yourself

Color type:
☐ Glossy White
☐ Glossy Black
☐ Matte White
☐ Matte Black
☐ Colors (12x12" square)
☐ Customer Supplied (\$3 foot)

*** No Rush Orders * 72 hours for full cut/weed/tape ***

Special Instructions / Notes

Vinyl / Paper Cutting Order Form 

Name: <input type="text"/>	Submitted Date: <input type="text"/>	Needed by Date: <input type="text"/>
Email: <input type="text"/>		
Phone: <input type="text"/>		

File Info

Folder File Name(s):

Illustration File:

Output Dimensions (L" x W"):

All Type outline?

☐ Paper Cutting

☐ Vinyl Cutting
 ☐ Weeding
 ☐ Taping
 ☐ Weed and tape yourself

Color type:
☐ Glossy White
☐ Glossy Black
☐ Matte White
☐ Matte Black
☐ Colors (12x12" square)
☐ Customer Supplied (\$3 foot)

*** No Rush Orders * 72 hours for full cut/weed/tape ***

Special Instructions / Notes

Digital Form Application: Development

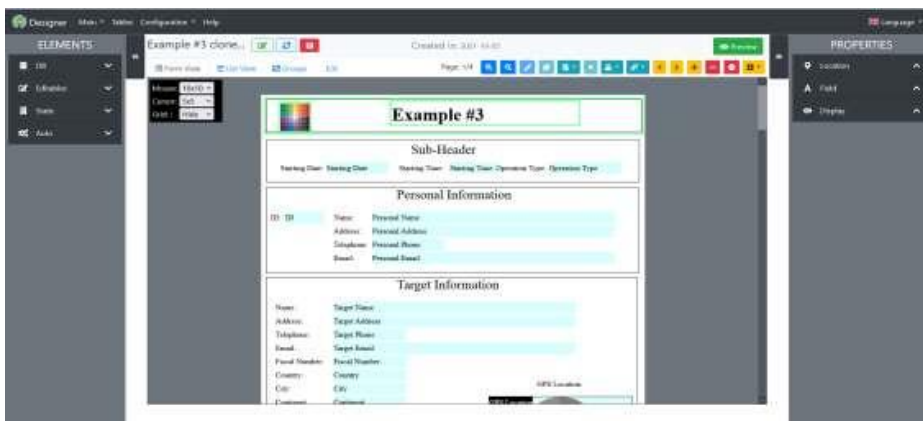
The developed solution is composed by seven main applications:

- Designer;
- Preview;
- Forms Manager;
- Operations Manager;
- Operation Editor;
- Inspectors App;
- Rest API;

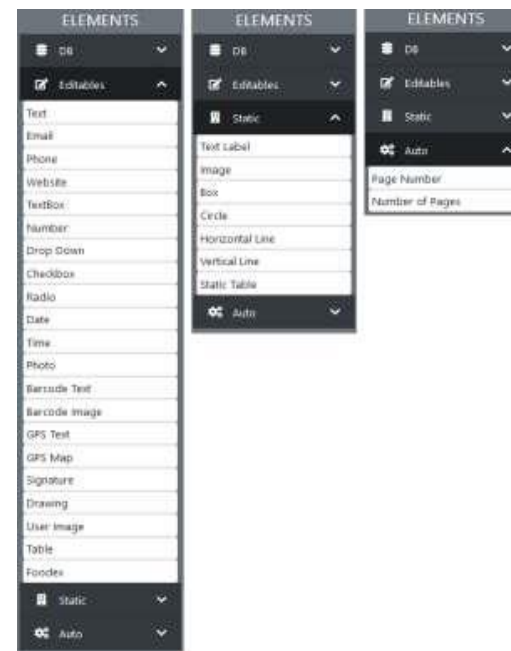


Digital Form Application: Designer

- The Designer or Editor is where a user can create new forms or edit existing forms:



The screenshot shows the IDRisk Designer interface. The main workspace displays a form titled "Example #3" with a sub-header "Sub-Header". The form contains two main sections: "Personal Information" and "Target Information". The "Personal Information" section includes fields for ID, Name, Address, Telephone, and Email. The "Target Information" section includes fields for Name, Address, Telephone, Email, Postal Number, Country, City, and Comment. The interface also features a left sidebar with "ELEMENTS" and a right sidebar with "PROPERTIES".

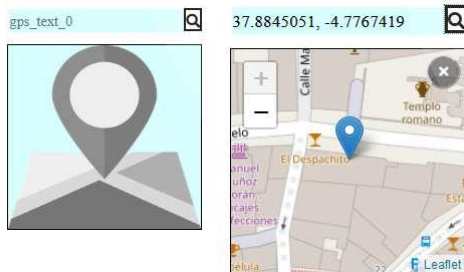


The image shows three panels of the "ELEMENTS" list in the IDRisk Designer interface. Each panel has a "DB" dropdown and a "Filter" button. The first panel shows a list of elements including Text, Email, Phone, Website, TextBox, Number, Drop Down, Checkbox, Radio, Date, Time, Photo, Barcode Text, Barcode Image, GPS Text, GPS Map, Signature, Drawing, User Image, Table, and Font. The second panel shows a list of elements including Text label, Image, Box, Circle, Horizontal Line, Vertical Line, and Static Table. The third panel shows a list of elements including Page Number and Number of Pages.

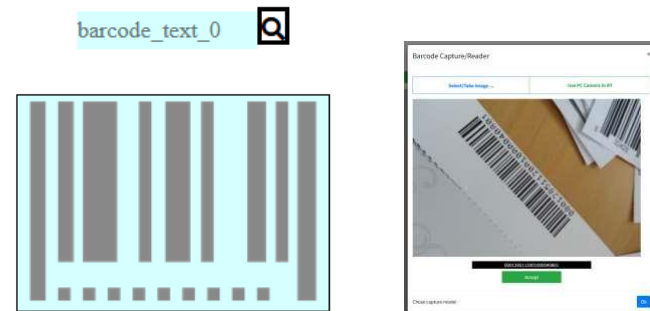
Digital Form Application: Designer

- Example of some elements:

Map (GPS)

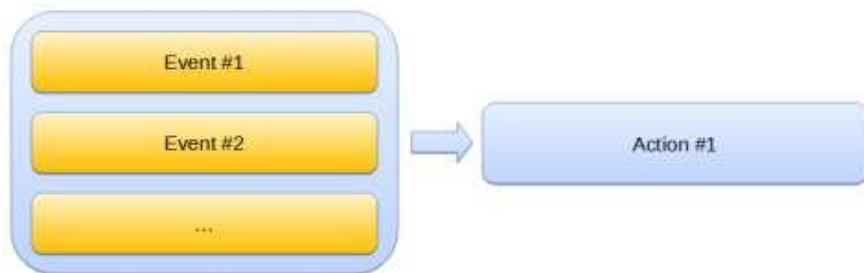


Barcode



Digital Form Application: Designer

- Events:
 - The Events/Actions systems allows the user to automate certain parts of the forms. One or more events are used to trigger a specific action:



Target Information

Name:	Target Name
Address:	Target Address
Telephone:	Target Phone
Email:	Target Email
Fiscal Number:	Fiscal Number
Country:	Country
City:	City
Continent:	Continent
Country Code:	Country Code
Official Language:	Official Language

GPS Location:



The form displays a list of fields for 'Target Information'. Each field has a label on the left and a corresponding input field on the right. The input fields are light blue. Below the list, there is a section for 'GPS Location' which includes a map icon with a location pin.

Digital Form Application: Preview

- Preview allows the user to visualize, test and print the current opened form. The form will be presented inside a tablet frame since it can help the designer to experience what the form's user will experience:



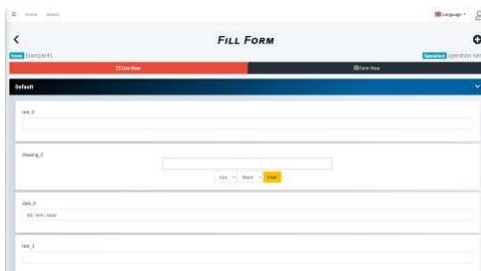
Digital Form Application: Inspector App

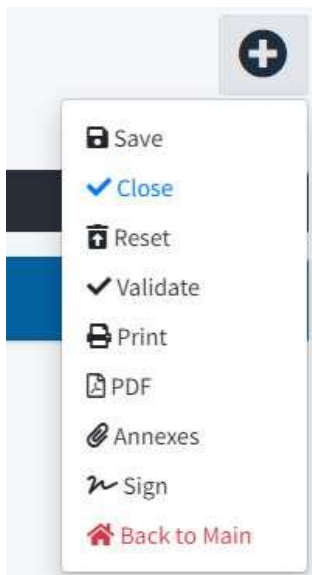
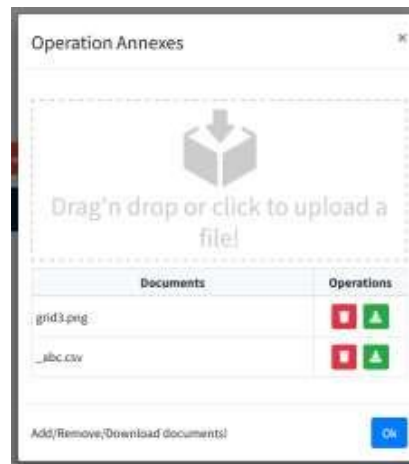
- This application is what a user uses to fill a form. An Operation corresponds to a form filled. This application, is a Progressive Web Application (PWA), which means it's designed to make the user feels like it's a native application and also, can operate without being connected to the web:



Digital Form Application: Inspector App

- Full of features:





Digital Form Application

Results

- System Available for download in <https://gitlab.com/arkhamlord666/forms>
- Support documents: <https://zenodo.org/record/6778397#.YxtTCXbMI2w>

FoodEx2: Exploratory Analysis for Automatic Classification

- The goal was to develop an automatic food solution using the FoodEx2 classification system that would assist the application user in classifying a sample based on its description.

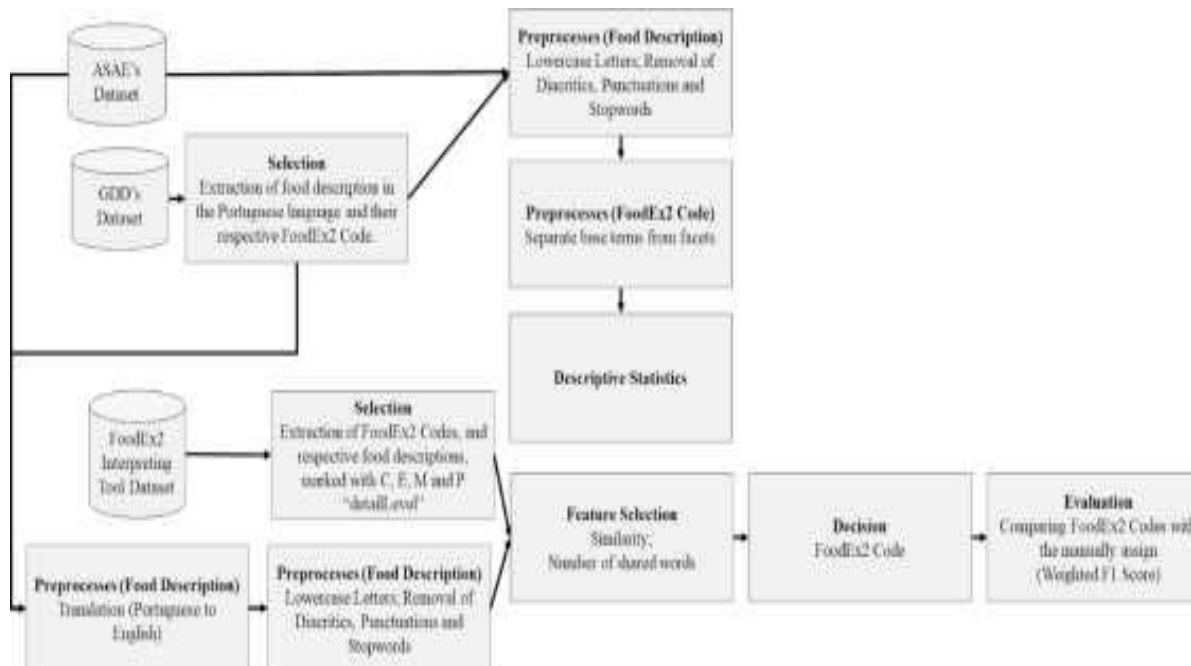
Food Description (PT) Ameixas descaroçadas, prontas a consumir, pasteurizadas, sem adição de conservantes
Food Description (EN) Pitted, ready-to-eat, pasteurized plums without preservatives

A01MB#F02.A068P\$F27.A01GQ\$F28.A07KG

FoodEx2: Exploratory Analysis for Automatic Classification

Approaches

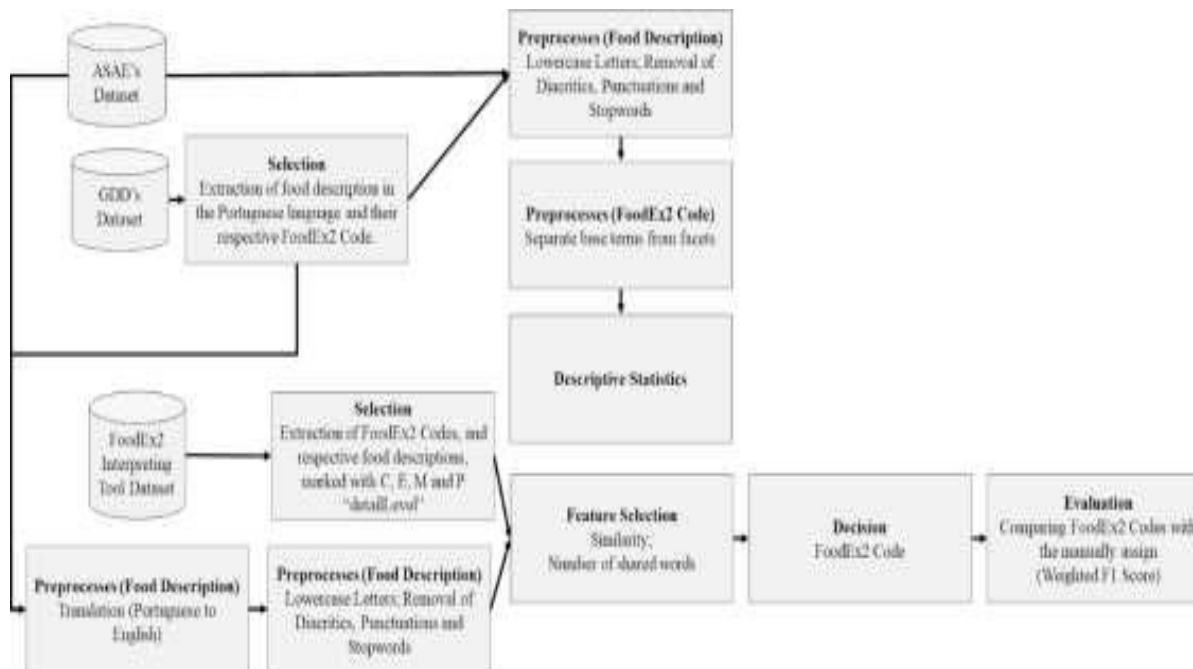
- Text similarity:
 - This method tries to assign a FoodEx2 code by matching a food description to one of the FoodEx2 code's standard descriptions. It works as a matching algorithm, but takes in consideration the Part-of-Speech (POS) tag of each word.



FoodEx2: Exploratory Analysis for Automatic Classification

Approaches

- Text similarity:
 - This method tries to assign a FoodEx2 code by matching a food description to one of the FoodEx2 code's standard descriptions. It works as a matching algorithm, but takes in consideration the Part-of-Speech (POS) tag of each word.



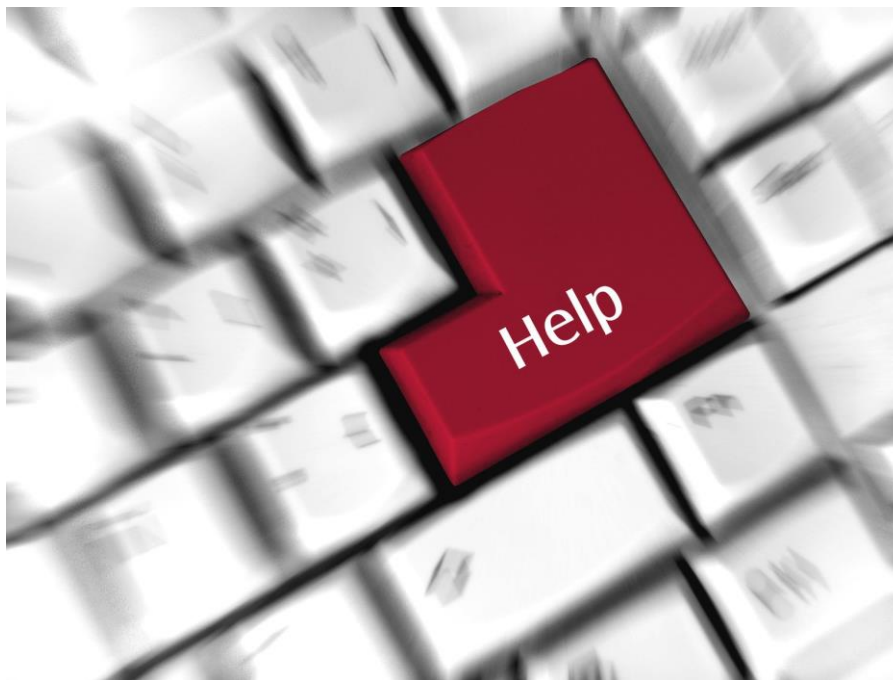
FoodEx2: Exploratory Analysis for Automatic Classification

Approaches

- Machine Learning for Natural Language Processing
 - Several algorithms were implemented and compared:
 - Bernoulli Naive Bayes;
 - Multinomial Naive Bayes;
 - Complement Naive Bayes;
 - K-Neighbors Classifier;
 - Decision Tree Classifier;
 - Random Forest Classifier;
 - Linear Support Vector Classification;
 - Logistic Regression;

FoodEx2

PT+HR datasets were not enough



FoodEx2

PT+HR datasets were not enough

- We asked for assistance:

- ✓ EFSA
- ✓ EFSA FP network
- ✓ EFSA Chemical Monitoring Data Network



New FoodEx2
datasets

FoodEx2

PT+HR datasets were not enough

- Met with other groups who were working to solve the same issue:

✓ EFSA, Sweden, France, Cyprus



FoodEx2

Results

- Datasets:
 - Compiled datasets (ASAE) + Found Online (GDD):
 - There are 78154 valid entries in the first dataset (ASAE dataset) and 700543 on the second (GDD);
- Text similarity:
 - The proposed algorithm obtained a low performance, by not being able to attribute the same FoodEx2 base code, to a food description, as a field worker. The translation might have negatively impact on the results.

Dataset	Nº of Objects	Nº of assigned codes	Nº of assigned codes that are equal to the originals	Weighted F1-Score
ASAE	78 157	51 938 (64%)	15 104 (19%)	0.18
GDD	700 543	538 709 (77%)	205 129 (29%)	0.28

FoodEx2

Results

- Datasets:
 - Compiled datasets (ASAE) + Found Online (GDD):
 - There are 78154 valid entries in the first dataset (ASAE dataset) and 700543 on the second (GDD);
- Machine learning algorithms:

Algorithms	Accuracy	F1-score (macro)
BernoulliNB	0.085	0.117
MultinomialNB	0.102	0.122
ComplementNB	0.090	0.104
LinearSVC	0.092	0.108
KNeighborsClassifier	0.038	0.058
LogisticRegression	0.099	0.122

The best results were only of 10% of accuracy...

FoodEx2 Results

- The results reflect the common problem that it's usually found in ML:

Not enough data!

- With the increase of interest in solving this problem, all of us could benefit of having access to datasets that could cover a wide range of information (food descriptions, language to language variations, etc.);
- The access to the shared datasets brings also the possibility of having better and better data quality over time, for future work.



Future perspectives

- Replicate ID Risk both in PT/HR and other countries that may be interested
- IDRISK 2.0
 - ✓ Further integration of SSD2 (directly integrate SSD2 catalogues, translation of SSD2 terms to country language)
 - ✓ **More data is indispensable**: allow EFSA data warehouse to be accessible for systems development
 - ✓ Work together and combine efforts with data/results exchange

***"if you want to go faster go alone,
but if you want to go far go together"***

African proverb



Thank you for your attention !