

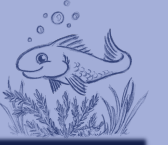
Katja Glass
Consulting

FAIR for Source Code

Digging the Gold of
Programming



Agenda



➤ Introduction

➤ The Use Case

➤ Opportunities

➤ Summary



Introduction



Typical Questions

- I did program this already! Where was it?
- Didn't we programmed something similar?
- There was a smart way, wasn't there?
- Someone programed that domain somewhere ...



Issue

- Similar programming multiple times
 - Thinking about issues again and again
 - Typically once-a-year issues
-



FAIR for Source Code

Introduction



F - Findable

- Find Programs, Patterns, Specifics
- Search by “Metadata”
 - Description
 - Author
 - Type
 - Titles (specific tables)
 - Input Datasets (specific domains)
 - Keywords

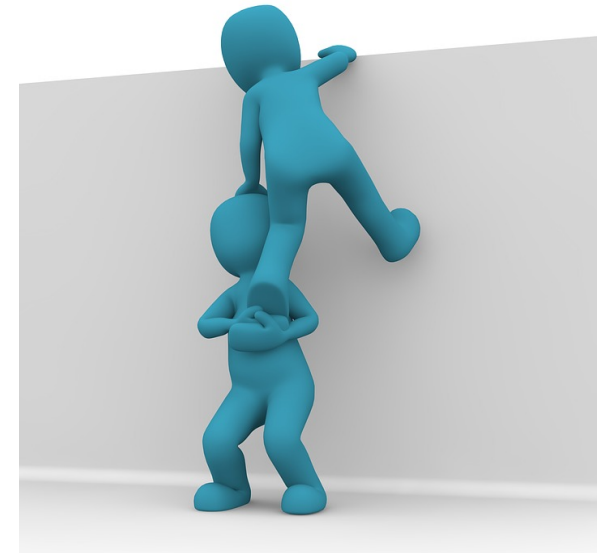


Introduction



A - Accessible

- Accessible location
 - Directory
 - URLs
 - GitHub
 - ...



Introduction



I - Interoperable

- Accessible by different tools
 - Metadata in common formats
 - Plain source code (not encrypted)
 - Ideally exchangeable (SAS <-> R <-> Python <-> ...)



Introduction



R- Re-Usable

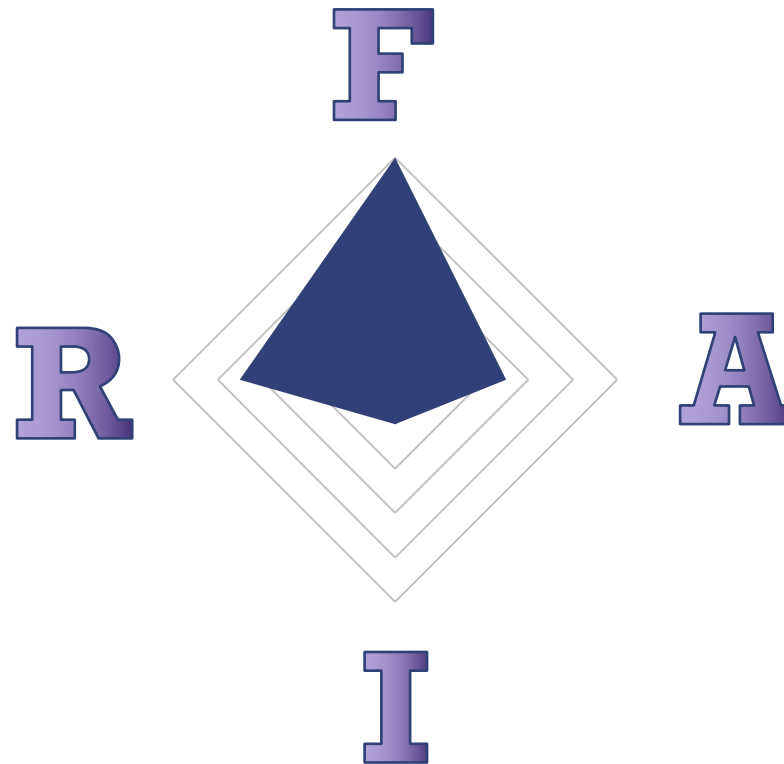
- Apply to different data/tools
 - Macros, Packages & Functions
 - Snippets
 - Patterns



Introduction

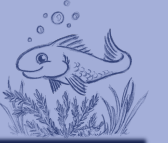


Importance of FAIR for Source Code



* assumptions

Agenda



- Introduction
 - **The Use Case**
 - Opportunities
 - Summary
-



The Use Case



Use Case

- Programs & Macros for various topics are available
- Making these FINDABLE!

- Open Source SAS Macros & Programs
 - PhUSE Whitepaper Scripts
 - SAS Macros by Scott Bass
 - SASjs Core Macros
 - ...

- Alias could be
 - Programs from different groups / departments
 - Programs/Macros from different people



The Use Case



- Findable ... but What?
 - File name? Not sufficient!



METADATA

The Use Case



Metadata

Create

- Using Hash-Tags like Twitter
(#INFILE, #CARDS, #HASH, #SQL-SORT, ...)
- Special Documentation
(Metadata file in general format YAML)

```
1  Keywords: AE, Severity
2  Script:
3    Desc   : Script description
4    Name   : Script name following the proposed
5    GCR    : r135
6    SRV    : 12345
7    Source : FDA JumpStart
8    Target : https://github.com/phuse-org/phuse-
9    Title  : AE Severity Analysis Panel
```

The Use Case



Metadata

Create

- Using Hash-Tags like Twitter
(#INFILE, #CARDS, #HASH, #SQL-SORT, ...)
- Special Documentation
(Metadata file in general format YAML)

```
1  Keywords: AE, Severity
2  Script:
3    Desc   : Script description
4    Name    : Script name following the proposed
5    GCR     : r135
6    SRV     : 12345
7    Source  : FDA JumpStart
8    Target  : https://github.com/phuse-org/phuse-
9    Title   : AE Severity Analysis Panel
```

Crawl

- Read & Analyze SAS Code
(INFILE statement, HASH statement, specific domain input dataset, Titles, ...)
- Read Header information

```
1  /*=====
2  Program Name      : IsNumD.sas
3  Purpose           : Checks if alphanumeric (charac
4                     valid numeric data (data step)
5                     Also sets additional flag vari
6                     IsInt, IsFloat, IsNonNeg, IsPo
7  SAS Version       : SAS 9.3
8  Input Data        : Character data in a data step
```

The Use Case



Metadata

- Must be maintained
 - Up-to-date
 - Reliable
-
- Ideally documentation close to source (header)
 - Crawling as automated metadata collection*



* Paper about crawling available in German (<https://www.glacon.eu/resources/2019-KSFE-Beitrag.pdf>)

The Use Case



Metadata Crawling

- Programming language can be anything (SAS, Python, R, Perl)

```
5 DATA allContent;  
6     ...  
7     ATTRIB parameter    FORMAT = $100. LABEL="Current Parameter";  
8     RETAIN name purpose sas_version input_data output_data mac_called  
9     RETAIN parameter file_lines;  
10  
11     INFILE "&scott_path/*.sas" FILENAME=fname EOV=eov END=end;  
12     INPUT;  
13     line = _INFILE_;  
14     filename = fname;  
15     ...  
16 RUN;
```

The Use Case



Metadata Format (CSV, XML, JSON)

- Various Attributes (Name, File, Location, Description, Author, Content, Date, ...)
- Different Attributes per File / Group

```
{  
  "SAS Macros by Scott Bass": {  
    "compare": {  
      "tool": "SAS Macros by Scott Bass",  
      "toolId": "SAS Macros by Scott Bass",  
      "name": "compare",  
      "filename": "compare.sas",  
      "description": "PROC COMPARE either...",  
      "type": "Macro",  
      "class": "Utility",  
      "language": "SAS",  
      "language_version": "9.1.3",  
      "author": "Scott Bass",  
      "creationDate": "2009-01-21",  
      ...  
    }  
  }  
}
```

The Use Case



Interface to Metadata

- Display & Search Metadata, but how?
 - Webpage
 - R-Shiny
 - SAS Interface
 - ...



The Use Case



Interface to Metadata

- Webpage*
 - Findable & Accessible by anyone
 - Intuitive



- React Framework
- Many others work as well



* <https://www.glacon.eu/portal>

The Use Case



Interface to Metadata

 Home Overview Programs Projects Info							
Tool	Name	Description	Type	Class	Language	Mod Date	Lineno.
Show All ▾			Show All ▾	Show All ▾	Show ▾		
FDA Jumpstart Scripts	ae_meddra_w_flag_gen	MedDRA at a Glance Analysis Panel - The MedDRA at a Glance Analysis Panel shows the user the adverse events that occurred in the study according to their place in the MedDRA hierarchy and allows them to compare any two arms according to a number of statistics.	Script	Tool	SAS	2011-05-08	653
FDA Jumpstart Scripts	ae_oncology_v1upd	AE Toxicity Analysis Panel - The AE Toxicity Analysis Panel provides the user with different analyses to help understand the adverse events present in the study and the frequency of their occurrence by treatment arm. Statistical and frequency reports are generated to provide the user with needed information.	Script	Tool	SAS	2011-06-02	408
FDA Jumpstart Scripts	ae_v1upd	AE Severity Analysis Panel	Script	Tool	SAS	2011-06-08	299
FDA Jumpstart Scripts	demographics_v1	Demographic Analysis Panel - The Demographic Analysis Panel provides the user with an overview of the number and percent of subjects that	Script	Tool	SAS	2011-06-02	1218

The Use Case



Interface to Metadata

ae_oncology_v1upd

FDA Jumpstart Scripts

Description:

AE Toxicity Analysis Panel - The AE Toxicity Analysis Panel provides the user with different analyses to help understand the adverse events present in the study and the frequency of their occurrence by treatment arm. Statistical and frequency reports are generated to provide the user with needed information.

More:

Find subject counts per arm for each adverse event broken down by toxicity grade Compare the AEs between two arms and find risk difference, relative risk, odds ratio, 95% confidence intervals, and p-value

File Name	ae_oncology_v1upd.sas
Type	Script
Class	Tool
Language	SAS
Version (Lang.)	9.2
Author	FDA/IBM
Last Update	2011-06-02
Input Data	AE, DM, EX
Input File	AE Toxicity Analysis.xls
Output File	Excel
Req. Macros	various
File Lines	408
Link Source	https://github.com/...
Link Source (Raw)	https://raw.githubusercontent.com/...
Link Specifications	https://github.com/...

The Use Case



Interface to Metadata (dynamic display)

File Name	demo_summary.sas
Type	Script
Keywords	Demographics, Table
Title	Demographic Summary
Script Type	table - summary
Stage	Contributed
Class	Standalone
Language	SAS
Version (Lang.)	9.2.1
File Lines	26
Link Source	https://github.com/...
Link Source (Raw)	https://raw.githubusercontent.com/...
Link Target	http://www.phusewiki.org/...

File Name	compare.sas
Type	Macro
Class	Utility
Language	SAS
Version (Lang.)	9.1.3
Author	Scott Bass
Last Update	2009-01-21
Input Data	Either Base and Compare datasets or Base and Compare libraries
Output Data	N/A
Req. Macros	parmv
File Lines	200
Link Source	https://github.com/...
Link Source (Raw)	https://raw.githubusercontent.com/...

The Use Case



- Dynamic interface
 - Additional Metadata defining structure
 - On new items -> update structure metadata

ae_oncology_v1upd
FDA Jumpstart Scripts

Description:
AE Toxicity Analysis Panel - The AE Toxicity Analysis Panel provides the user with different analyses to help understand the adverse events present in the study and the frequency of their occurrence by treatment arm. Statistical and frequency reports are generated to provide the user with needed information.

More:
Find subject counts per arm for each adverse event broken down by toxicity grade. Compare the AEs between two arms and find risk difference, relative risk, odds ratio, 95% confidence intervals, and p-value.

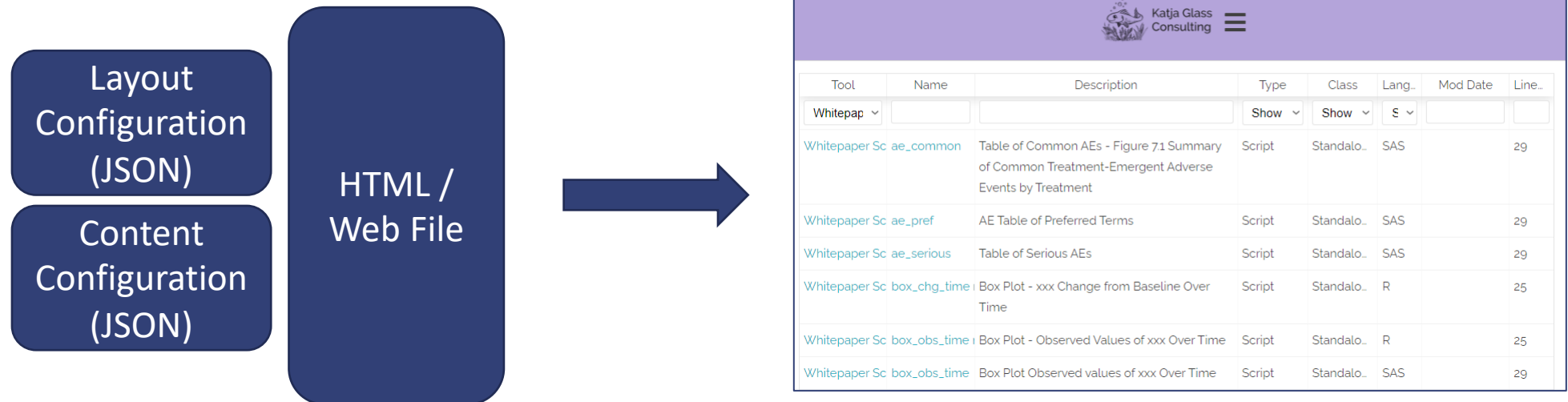
File Name	ae_oncology_v1upd.sas
Type	Script
Class	Tool
Language	SAS
Version (Lang)	9.2
Author	FDA/IBM
Last Update	2011-06-02
Input Data	AE, DM, EX
Input File	AE Toxicity Analysis.xls
Output File	Excel
Req. Macros	various
File Lines	408
Link Source	https://github.com/
Link Source (Raw)	https://raw.githubusercontent.com/
Link Specifications	https://github.com/

```
"description" : {  
  "displayDetailsView" : false,  
  "displayDetailsText" : true,  
  "type" : "string",  
  "label" : "Description"  
},  
"authors" : {  
  "displayDetailsView" : true,  
  "type" : "string",  
  "label" : "Author(s)"  
}
```

The Use Case



- Dynamic interface
 - Display whatever is in the data
 - One Webpage, different content
 - “Plug” different metadata



The Use Case

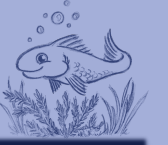


➤ The Use Case

- 1) Figure out what do you need
- 2) Collect corresponding metadata with crawling
- 3) Display results, configuration driven
- 4) Make is searchable (& accessible, reusable)



Agenda



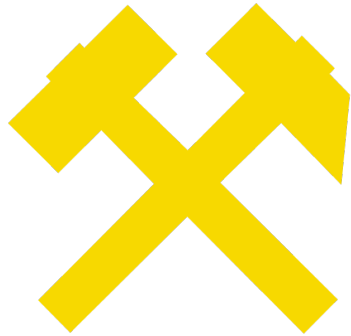
- Introduction
 - The Use Case
 - **Opportunities**
 - Summary
-



Opportunities



- Entrance to your programs!
- Digging the gold!
- Effort is low, benefit high!



Opportunities



- Has this table been programmed?
 - Crawl your programs for titles
 - Make them searchable

```
156  options
157  title1 j=1 "Table X: Dispostion Table";
158  title2 j=1 "##";
159  title3 j=1 "Subject Disposition";
160  footnote1 j=1 "Abbreviations: N=total number of
```

Opportunities



➤ Catalog of Table Programs



Opportunities



- Has this domain been used somewhere?
 - Crawl your programs for specific datasets (SDTM.DI)
 - Make them searchable

4	
5	DATA SDTM.DI;
6	SET IN.devices;
7	...
8	RUN ;

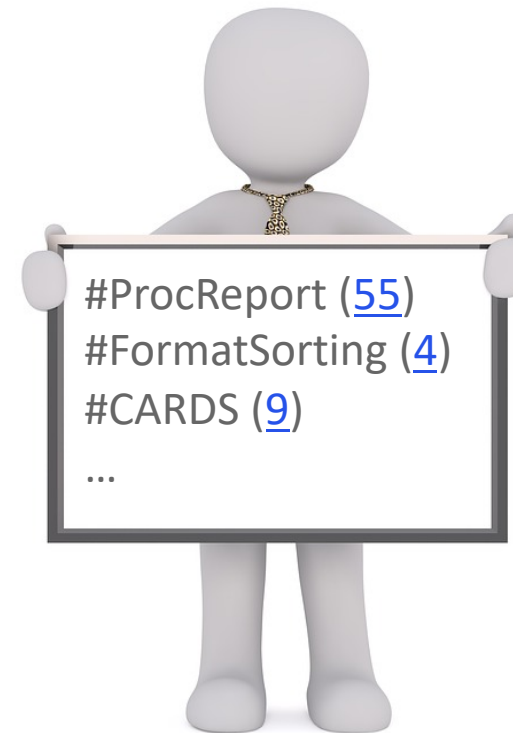
- Catalog of special domain usages

SDTM.DI (2)
SDTM.DT (5)
SDTM.PK (27)
...

Opportunities



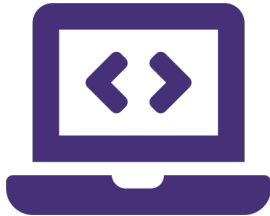
- Hashtag knowledge repository
 - Create “Hashtags” for “important to remember files”
 - #hash #hash-sort
 - #CARDS #Infile #Input
 - #ProcReport
 - #FormatSorting



Opportunities



- Personal knowledge repository
 - Create a “generic” Webpage
 - Colleagues can “plugin” their metadata & configuration
 - Highly customizable interface to your programs!



Standard programs



Table related
programs



Katjas programs



Pauls programs

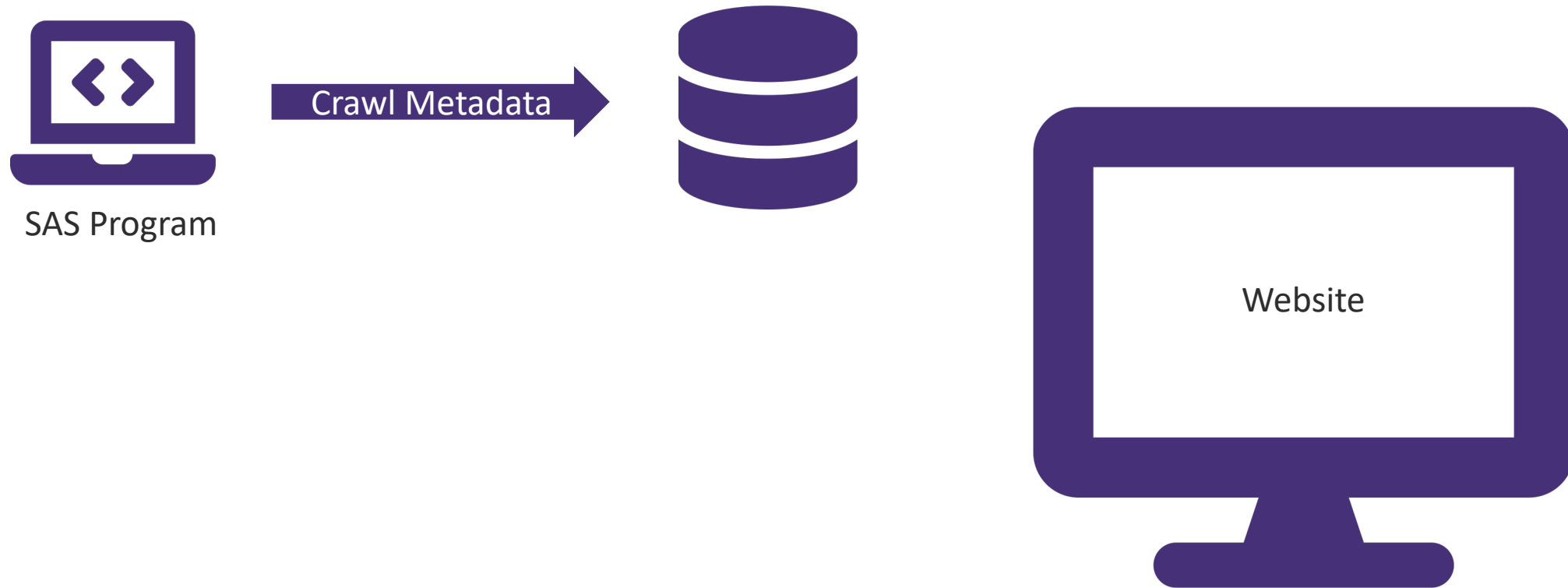


Domain related
programs

Opportunities



➤ Personal knowledge repository



Opportunities



➤ Personal knowledge repository

Configuration

Enter metadata file ...

Load

Item	Overview	Detail View
Name	☒	☒
Description	☒	☒
Author	☐	☒
Keywords	☒	☒

Opportunities



➤ Personal knowledge repository

Browse Programs		
Search overall metadata		
Name	Description	Keyword
compare	compares two datasets	checks
derive_date	derive date in ISO format	mapping
print_quick	prints a dataset for quick review	report , check
...	...	...

Opportunities

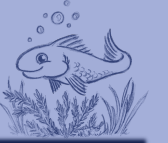


- Ideal Open Source Project!!!
 - Small scale project
 - Generic
 - No secrets (header structure)
 - Corporate designs through styles

 - Ideal for upskilling
 - Project for students
-



Agenda



- Introduction
- The Use Case
- Opportunities
- **Summary**



Summary



- FAIR Source Code
 - Easy accessible programming knowledge repository
 - Personal
 - Corporate
 - Worldwide
 - Findable most important
 - Implementation is easy
 - Perfect Open Source fit



Summary



- FAIR Source Code
 - After **F** indable
- Consider other aspects
 - Accessible
 - (Interchangeable)
 - Re-usable





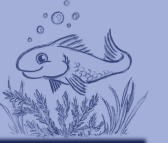
Open Source

FAIR

Solutions



Questions?



Katja Glass Consulting

info@glacon.eu

www.glacon.eu

www.glacon.eu/portal
