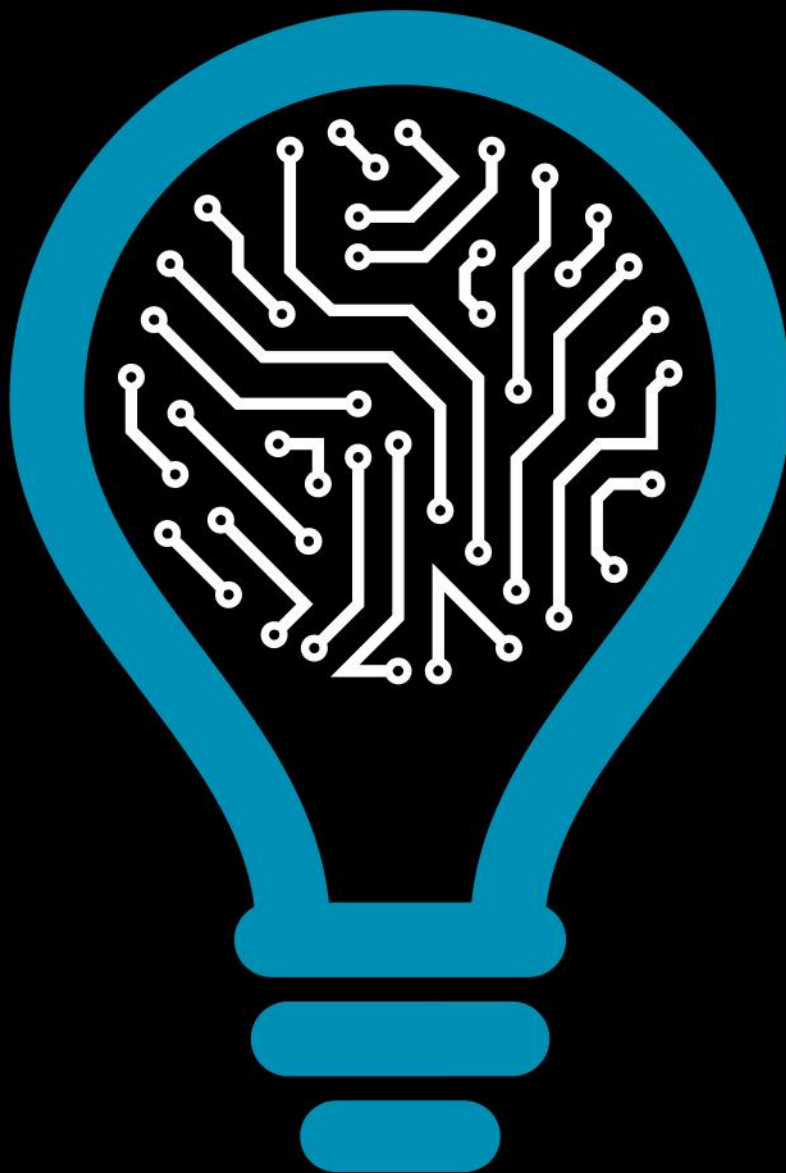


Da produção
de conhecimento
à inovação de
base científica



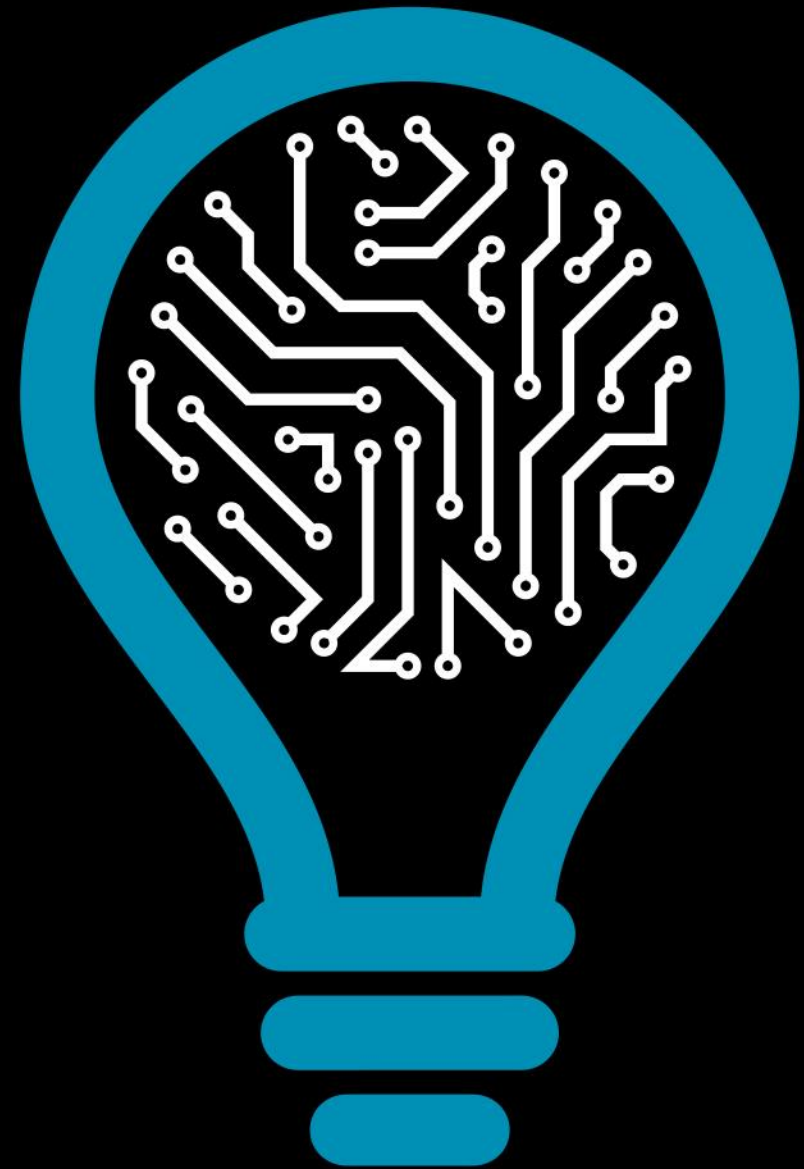
**INSTITUTO DE ENGENHARIA
DE SISTEMAS E COMPUTADORES,
TECNOLOGIA E CIÊNCIA**

Um repositório baseado no paradigma de base de dados

Artur Rocha, Susana Barbosa, Cristina Ribeiro, Gabriel David

2º Fórum Gestão de Dados de Investigação

31 de março de 2017



1 Paradigma ficheiros

2 Paradigma base de dados

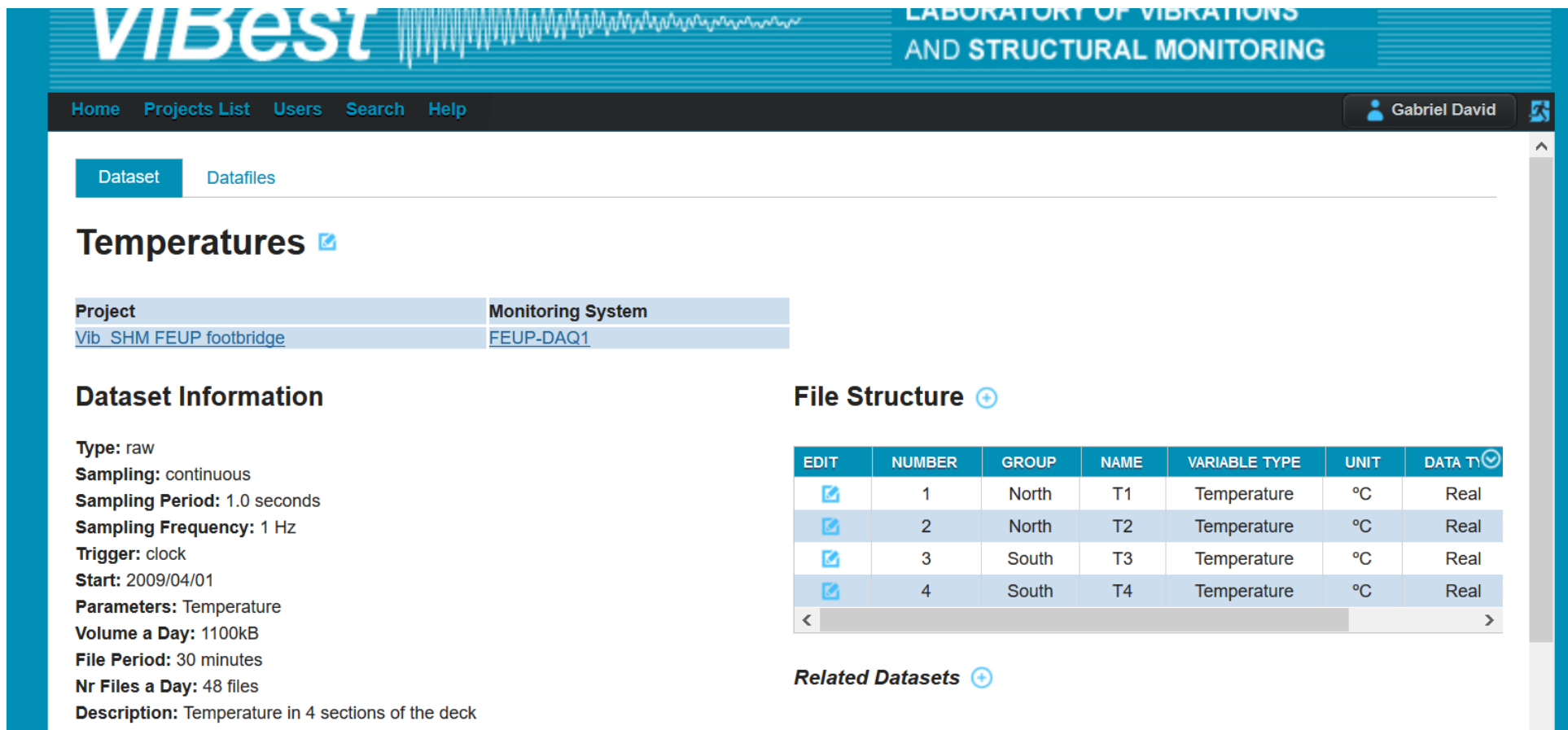
Paradigma ficheiros

Paradigma ficheiros

- Dados organizados em ficheiros
- Formatos habitualmente simples (csv, tab) com ou sem cabeçalho
- Muitas observações
- Fáceis de processar linearmente
- Metadados
 - ficheiro específico
 - BD do domínio
 - Repositórios genéricos

	1	2	3	4
1	8.367600	7.845312	7.549594	8.330053
2	8.367837	7.846104	7.548247	8.328072
3	8.366570	7.847055	7.547455	8.328944
4	8.366253	7.844758	7.548643	8.329815
5	8.368708	7.844203	7.546188	8.328469
6	8.366253	7.844045	7.548485	8.329894
7	8.366966	7.845154	7.546268	8.326726
8	8.365936	7.845550	7.548485	8.329498
9	8.368154	7.843332	7.548564	8.326488
10	8.365857	7.843887	7.545001	8.327756
11	8.363718	7.844520	7.547614	8.327201
12	8.364748	7.843570	7.544842	8.326092
13	8.363797	7.842065	7.548010	8.327518
14	8.365857	7.843095	7.545396	8.325855
15	8.363322	7.843887	7.545792	8.327280
16	8.365282	7.842226	7.545051	8.326488

Metadados em base de dados especializada



VIBest LABORATORY OF VIBRATIONS AND STRUCTURAL MONITORING

Home Projects List Users Search Help Gabriel David

Dataset Datafiles

Temperatures

Project	Monitoring System
Vib SHM FEUP footbridge	FEUP-DAQ1

Dataset Information

Type: raw
 Sampling: continuous
 Sampling Period: 1.0 seconds
 Sampling Frequency: 1 Hz
 Trigger: clock
 Start: 2009/04/01
 Parameters: Temperature
 Volume a Day: 1100kB
 File Period: 30 minutes
 Nr Files a Day: 48 files
 Description: Temperature in 4 sections of the deck

File Structure

EDIT	NUMBER	GROUP	NAME	VARIABLE TYPE	UNIT	DATA TYPE
<input checked="" type="checkbox"/>	1	North	T1	Temperature	°C	Real
<input checked="" type="checkbox"/>	2	North	T2	Temperature	°C	Real
<input checked="" type="checkbox"/>	3	South	T3	Temperature	°C	Real
<input checked="" type="checkbox"/>	4	South	T4	Temperature	°C	Real

Related Datasets

Vantagens dos repositórios de domínio

- Maior riqueza dos parâmetros
- Metadados de estrutura
- Metadados com descrição dos equipamentos e condições da experiência
- Metadados de contexto (projeto, equipa, organização)
- Serviços mais elaborados
- Requer manutenção

Repositório genérico

CKAN

Welcome to the INESC TEC research data repository.

This data repository showcases datasets produced or used by INESC TEC researchers and their partners. It is an embodiment of our institutional commitment to Open Data in research.



Search data

E.g. environment



Popular tags

recommender systems

radon

playlists

Metadados genéricos

- Formatos mais ou menos standard (Dublin Core)
- Metadados de contexto em ficheiro auxiliar
- Menos dependente de manutenção
- Múltiplos domínios científicos
- Não adequada para acessos arbitrários ao conteúdo
- Depósito mediado por curador

The screenshot shows a dataset page with a navigation bar containing 'Dataset', 'Groups', and 'Activity Stream'. The main title is 'Albergaria Fire Fighters ECG 2010'. Below the title is a description: 'Monitoring of Albergaria Firefighters ECG during emergency events during the summer of 2010 and correspondent event labels.' There is a section for 'Data and Resources' with a file named 'SESSION-2010-06-07-01-OUTSIDE.zip' and a description 'Approximately 59 hours of ECG missions recordings of 5 Firefighters in .bdd...'. An 'Explore' button is visible. Below this is a search bar containing 'AlbergariaFirefighters'. The 'Additional Info' section contains a table with the following data:

Field	Value
Author	Joana Paiva, João Paulo Cunha
Last Updated	1 March 2017, 15:31 (UTC+00:00)
Created	19 January 2017, 13:23 (UTC+00:00)
dc.Contributor	Vital Responder 1.0 team, VR2Market team
dc.Coverage.Spatial	Albergaria
dc.Coverage.Temporal	February-July 2010
dc.Date	Summer, 2010
dc.Format	*.bdd, *.xls

Paradigma Base de datos

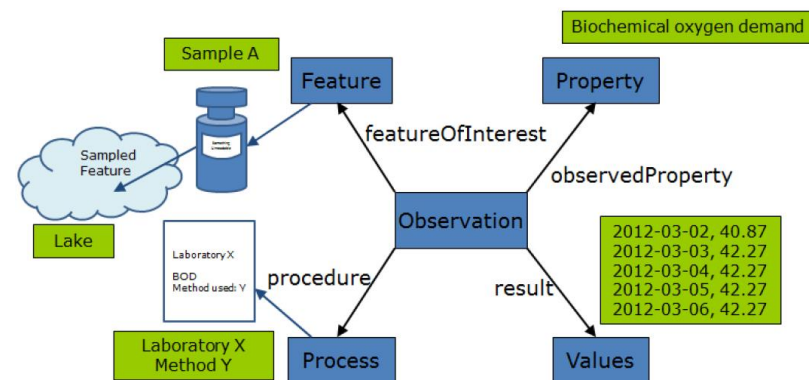
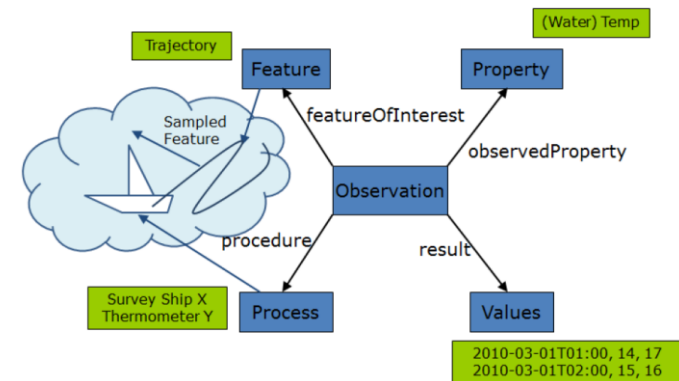
Paradigma base de dados

- Cada *observação* corresponde a um registo numa base de dados
- Permite
 - Relacionamento de dados
 - Pesquisa em SQL
 - Montagem de serviços elaborados
 - Metadados sofisticados na mesma plataforma dos dados
- Projetos com dados georreferenciados
 - Interoperável → Diretiva INSPIRE
 - Modelo baseado no Sensor Observation Service do OGC



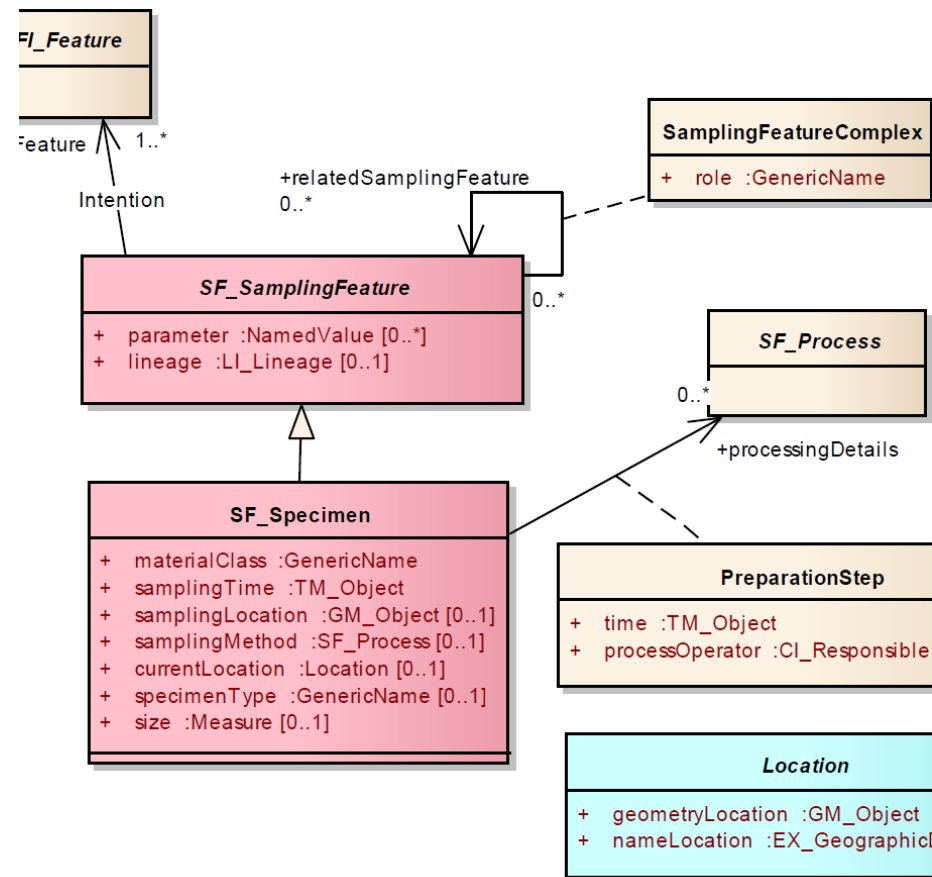
Modelo de observação

- Feature of Interest (feature geográfica)
- Procedimento (semiestruturado)
- Observação
- Propriedade observada
- Valor
- Generalidade dos conceitos pode levar a implementações lentas



Suporte a protocolos complexos

- Capacidade de relacionar etapas de recolha de dados em campanhas com processamentos laboratoriais subsequentes
- Conceito de amostra





Metadados de contexto

- Projetos
- Campanhas
- Equipamentos
- Equipas

SEA BIO DATA Administration Campaign Observations Samples Summary Fishing effort Administrator (BIOMETORE7)

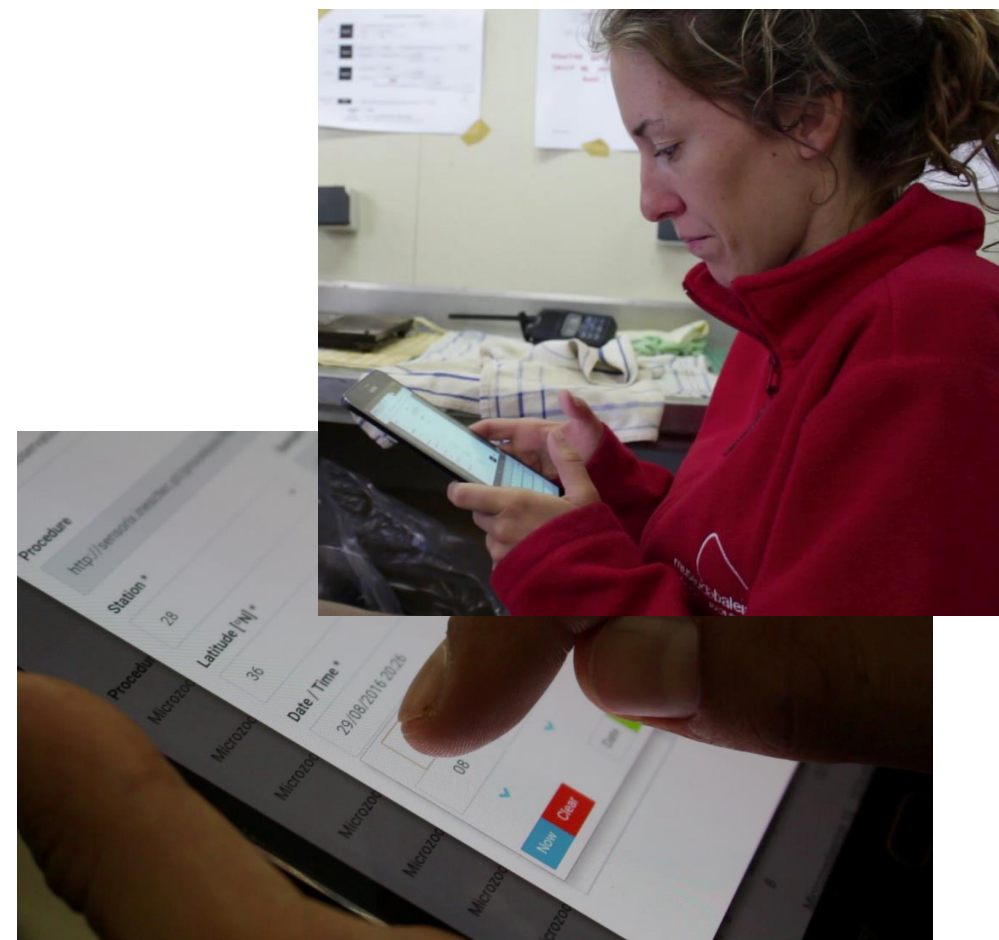
Campaigns > BIOMETORE7

Campaign BIOMETORE7
 Ship NI Noruega
 From 2016-08-16
 To 2016-09-03
 Chief Antonina dos Santos

Station	Latitude	Longitude	Depth	BONGO60	BONGO90	CTD	Microzooplankton	Location
5	37.2763055555556	-10.2382277777778	3827	7	7	3	3	Gorringe View Station Details
6	37.1197638888889	-10.4707472222222	3309	7	7	3	3	Gorringe View Station Details
7	36.9646583333333	-10.7273111111111	2290	7	7	3	3	Gorringe View Station Details
8	36.8708416666667	-10.8950666666667	1429	7	7	3	3	Gorringe View Station Details
9	36.8225361111111	-10.9837583333333	923	7	7	3	3	Gorringe View Station Details

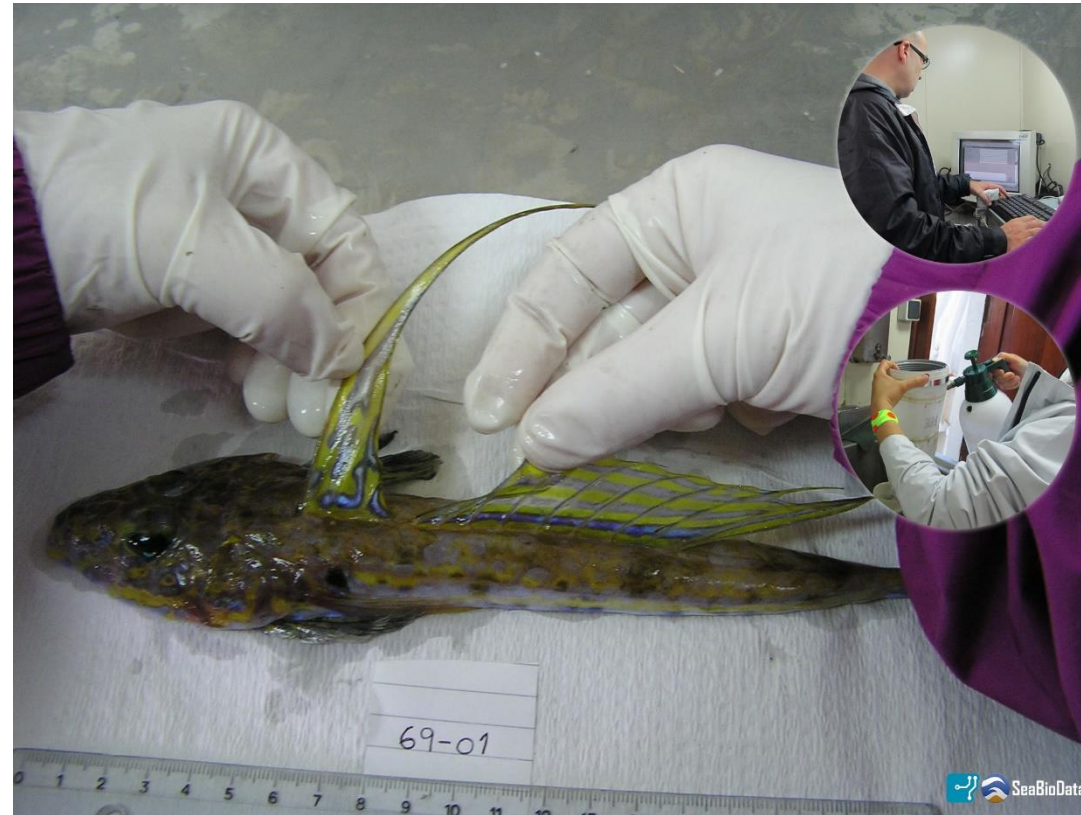
Ferramentas de recolha em tablets

- Utilização de tablets como caderno de registo
- Formulários configuráveis que produzem as observações no formato pretendido
- Ajudas ao registo com campos de vocabulário controlado, adição automática de elementos como as coordenadas GPS ou o instante



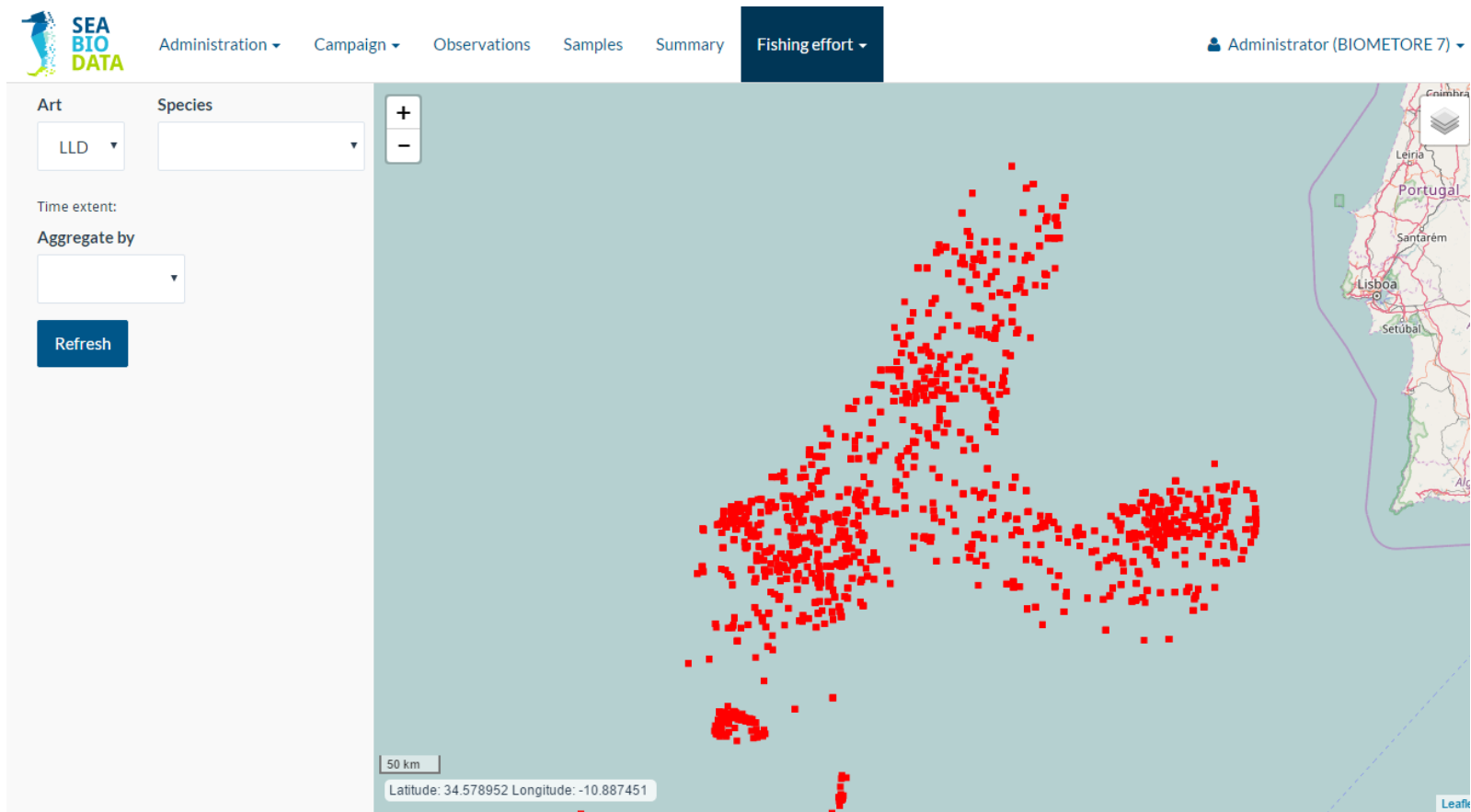
Documentos anexos

- Fotografias
- Vídeos
- Sons



- Repositório dos ficheiros originais para efeito de documentação

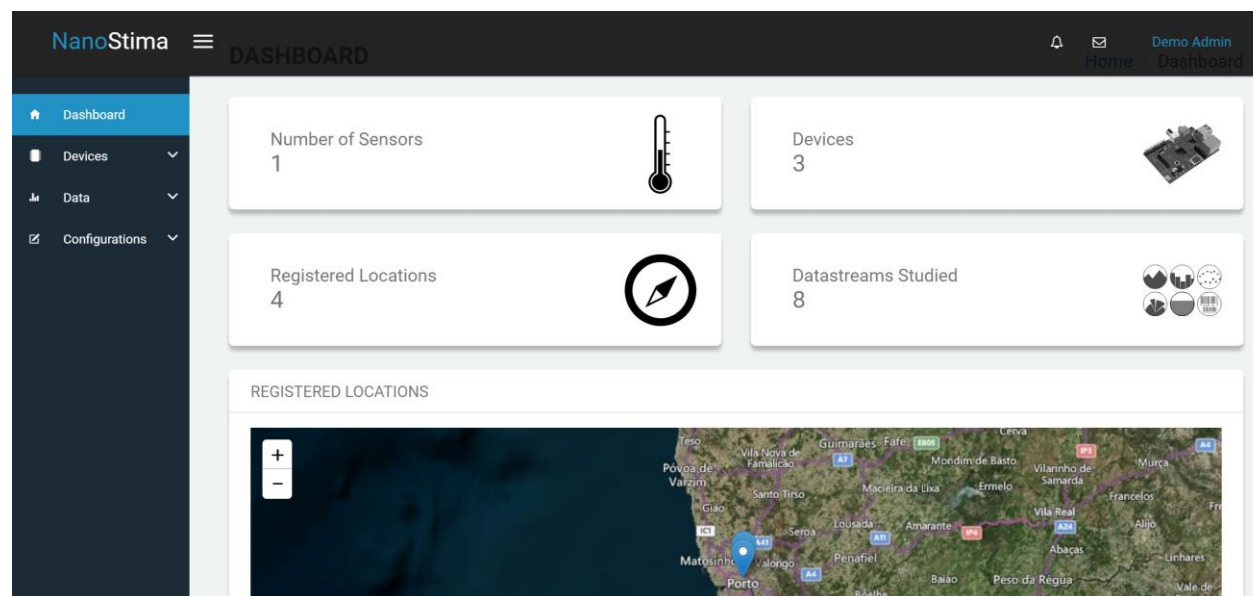
Montagem de serviços de exploração






Dados em tempo real - IoT

- Modelo comum para “streams” de dados
- Baseado na norma SensorThings do OGC
- RESTful API
- JSON encoded





Consumível por clientes genéricos

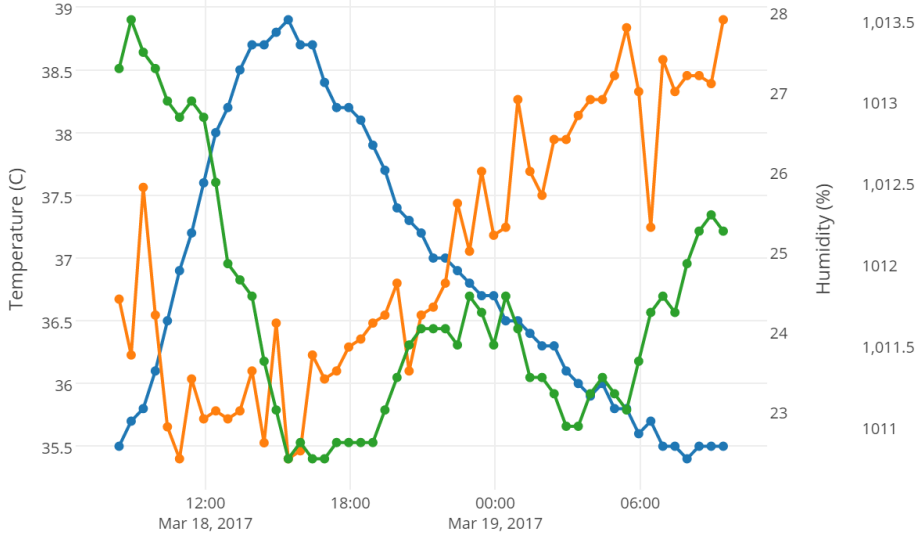
SensorThings Share


Templates ▾
Tools ▾
Share ▾
Run

```

1 <!DOCTYPE html>
2 <html>
3 <head>
4   <meta charset='utf-8' />
5   <meta name='viewport' content='initial-scale=1,maximum-scale=1,user-scalable=no' />
6   <link href='https://fonts.googleapis.com/css?family=Open+Sans:400,300,700' rel='stylesheet' />
7   <script src='https://cdn.plot.ly/plotly-1.22.0.min.js'></script>
8   <script src='http://sdk.sensorup.com/sta-chart/sta-chart-0.0.8.min.js'></script>
9 </head>
10 <body>
11   <div style='font-family: 'Open Sans', sans-serif, Arial; font-size: 12px;'>
12   </div>
13 </body>
14 </html>
15 </head>
16 <body>
17 <div id="vis"></div>
18 <script type="text/javascript">
19   stachart.generateChart({
20     'staBaseUrl': 'http://nanostima-rl3-2-2.inescn.pt:8080/sensorthings/noauth/v1.0',
21     'datastreamIds': [ '8', '9', '10' ],
22     'observations': {
23       retrieveAll: true,
24       timeRangeCode: 'last',
25       timeRangeAmount: 3,
26       timeRangeUnit: 'd'
27     },
28     'titleTemplate': 'Comparison',
29     'plotlyData': [
30       { name: 'Temperature', yaxis: 'y' },
31       { name: 'Humidity', yaxis: 'y2' },
32       { name: 'Pressure', yaxis: 'y3' }
33     ],
34   });

```



The chart displays three data series over a 24-hour period. The left y-axis represents Temperature (C) from 35.5 to 39. The right y-axis represents Humidity (%) from 23 to 28 and Pressure (mbar) from 1,011 to 1,013.5. The x-axis shows time from 12:00 on Mar 18, 2017 to 06:00 on Mar 19, 2017. The Temperature series (blue) peaks at approximately 38.8°C around 18:00. The Humidity series (green) peaks at approximately 27.5% around 06:00. The Pressure series (orange) peaks at approximately 1,013.5 mbar around 06:00.

Conclusão

- Os repositórios com dados em BD justificam-se quando os protocolos de recolha e análise dos dados envolvem várias fases
- Há a perspetiva de se fazerem processamentos relacionando diversos datasets
- Se pretendem pesquisas finas, usando dados georreferenciados
- A generalidade dos modelos requer implementações cuidadosas para evitar peso excessivo

Conversão entre paradigmas

Exportação de

conjuntos de observações selecionadas

para

datasets em ficheiros

utilizáveis

em troca de dados

ou em armazenamento em repositórios no outro paradigma

Trabalho em curso

- Atribuição de DOI para elevar a robustez das citações
- Como citar séries temporais em aberto?
- Como estender o modelo das observações para incluir os metadados de contexto de forma normalizada, preparando uma integração na rede de dados abertos ligados?
- Mais serviços de visualização