

*Vers une infrastructure numérique
nationale pour les SP :
Le projet « 65 Millions
d'Observateurs »*

Yvan Le Bras Marie Delannoy Marianne Linares Grégoire Loïs Romain Julliard

CESCO UMR 7204
65 Millions d'observateurs

Projet

Station de biologie marine de Concarneau

Forces et faiblesses des sciences participatives



- **un succès scientifique :**
 - plus de 80 publications à partir de données issues de Vigie-Nature
 - production d'indicateurs
- **un succès public :** plus de 15 000 observateurs actifs chaque année



- **Epuisement de l'animation**
- **Erosion de la participation**

Constats : Des freins liés aux outils :

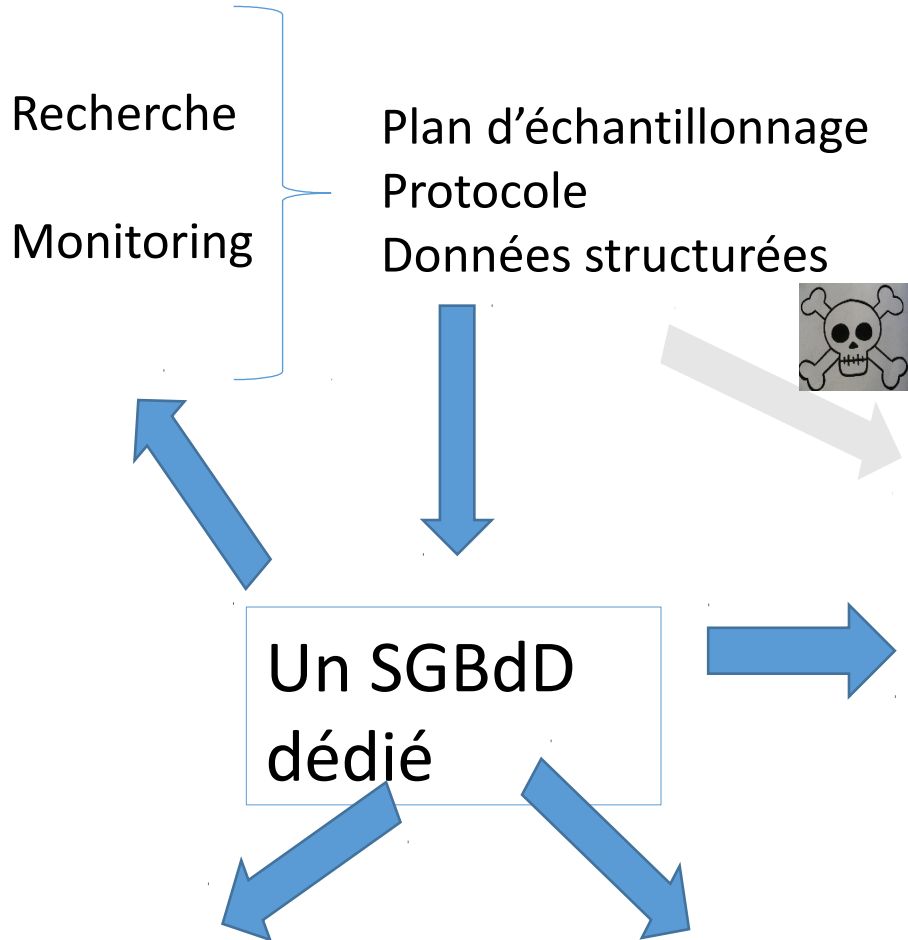
- **l'interopérabilité des données**
- **l'animation des programmes**
- **la création et la maintenance des outils**

1. Une difficulté : stocker des données structurées et protocoles

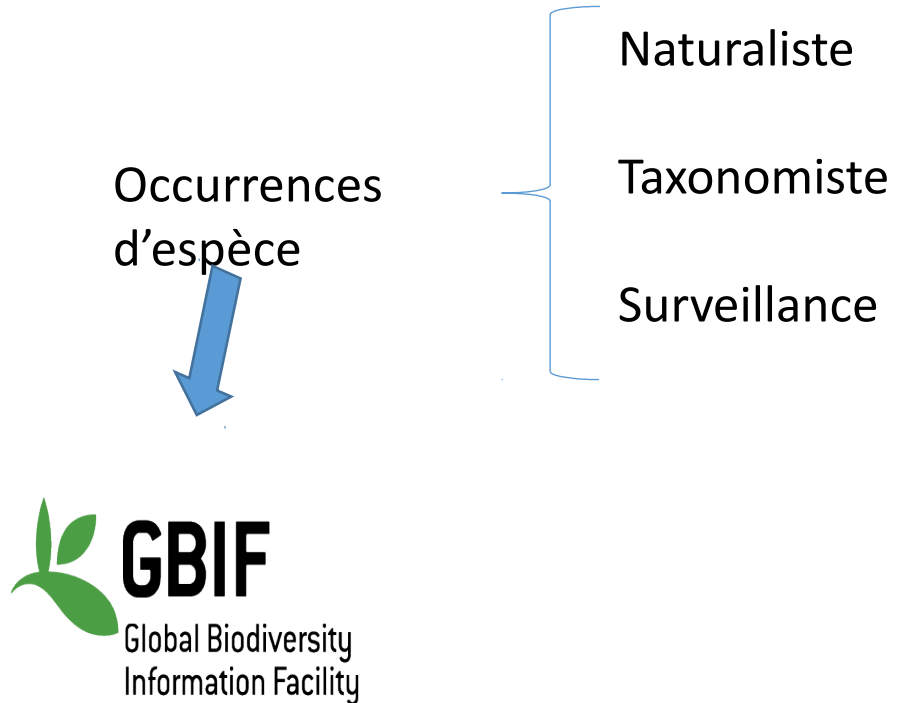
**=> Développement d'un
système de gestion de base
de données pour les sciences
participatives**

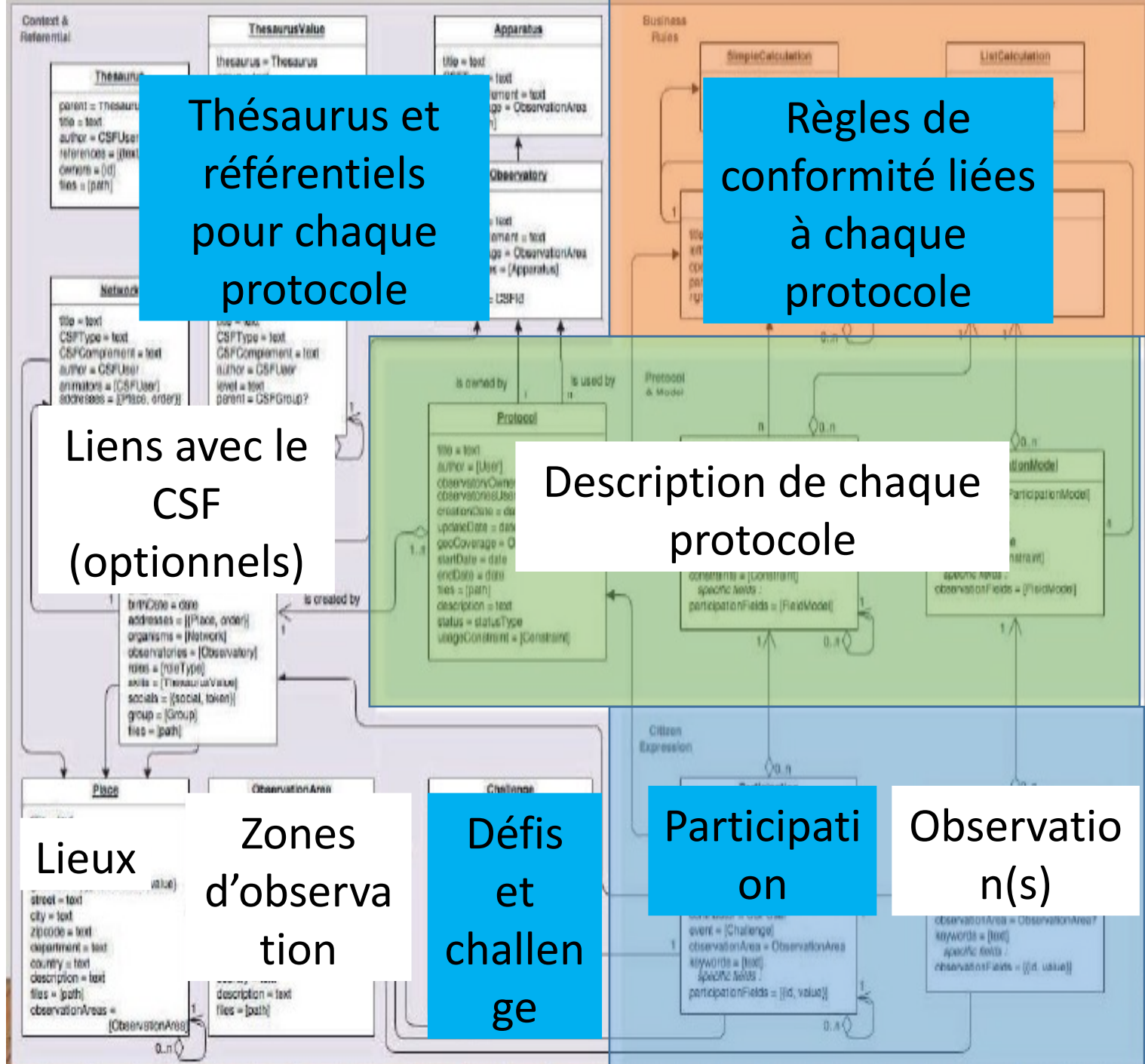
Données sur la nature : concilier deux logiques

Ecologie



Biodiversité





Thésaurus et référentiels pour chaque protocole

Règles de conformité liées à chaque protocole

Liens avec le CSF (optionnels)

Description de chaque protocole

Lieux
street = text
city = text
zip code = text
department = text
country = text
description = text
files = [path]
observationAreas = [ObservationArea]

Zones d'observation
description = text
files = [path]

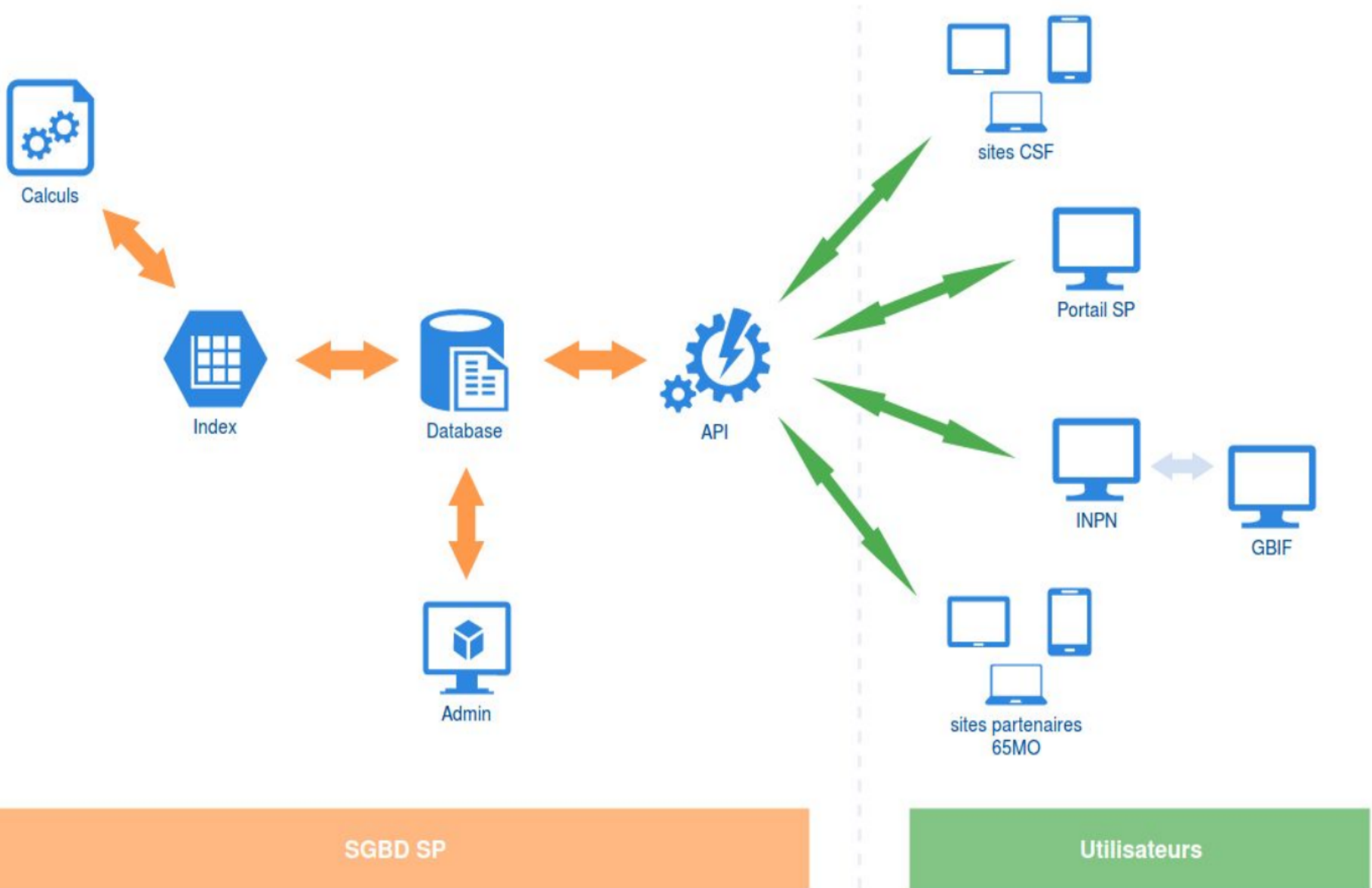
Défis et challenge

Participation
event = [Challenge]
observationArea = ObservationArea
keywords = [text]
specificFields = [field, value]
participationFields = [id, value]

Observation(s)
observationArea = ObservationArea?
keywords = [text]
specificFields = [field, value]
observationFields = [id, value]



écosystème du SGBD SP



Les services rendus par le SGBdD

1. Qualité et normalisation des données

- Définition des protocoles
- Cohérence et intégrité (règles métiers, techniques et de cohérence)

2. Stockage des données (*duplicata*)

- Sécurisation

3. Circulation des données :

- Au sein d'un observatoire, avec les relais locaux
- Entre observatoires
- Vers des bases de données extérieures (GBIF, INPN....)

4. Outils de traitement des données

- Capacités de calcul et d'analyse
- Représentations graphiques et spatiales

2. Un retour d'expérience : importance de la plateforme sociale pour la qualité des données



SPIPOLL

**Suivi Photographique
des Insectes POLLinisateurs**





AFFICHER

2010-09-17

Nom de la Fleur: Taxon inconnu de la clé (aster)

Il s'agit d'une fleur : plantée

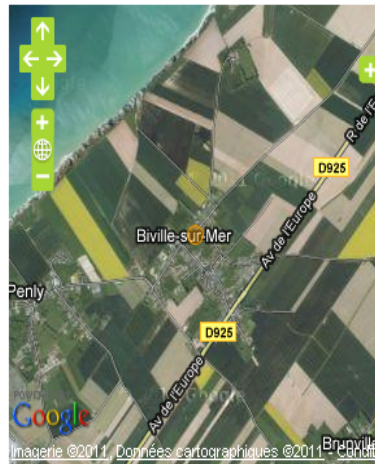
Il s'agit d'une habitat : jardin privé /

BIVILLE-SUR-MER (76098), SEINE-MARITIME (76),

HAUTE-NORMANDIE (23)

par : étamines

[TOUTES SES COLLECTIONS DANS LES GALERIES](#)



Dernière(s) identification(s):
Les Mouches à damier

AFFICHER



Dernière(s) identification(s):
La Syrritte piaulante

AFFICHER



Dernière(s) identification(s):
Les Mégachiles rayés

AFFICHER



Dernière(s) identification(s):
Les Mouches aux reflets métalliques

AFFICHER

COMMENTAIRES DES INTERNAUTES

par : cybelle 2011-04-23

Superbe votre collection Etamines, les insectes se bouscuaient pour vous faire plaisir, vous avez dû vous régaler.

par : Fernand 2011-04-23

Très belles photos. Les fleurs sont magnifiques et les insectes également. Beau travail étamines .

2. Un retour d'expérience : importance de la plateforme sociale pour la qualité des données

- Conseil**
- Imitation**
- Contrôle social**

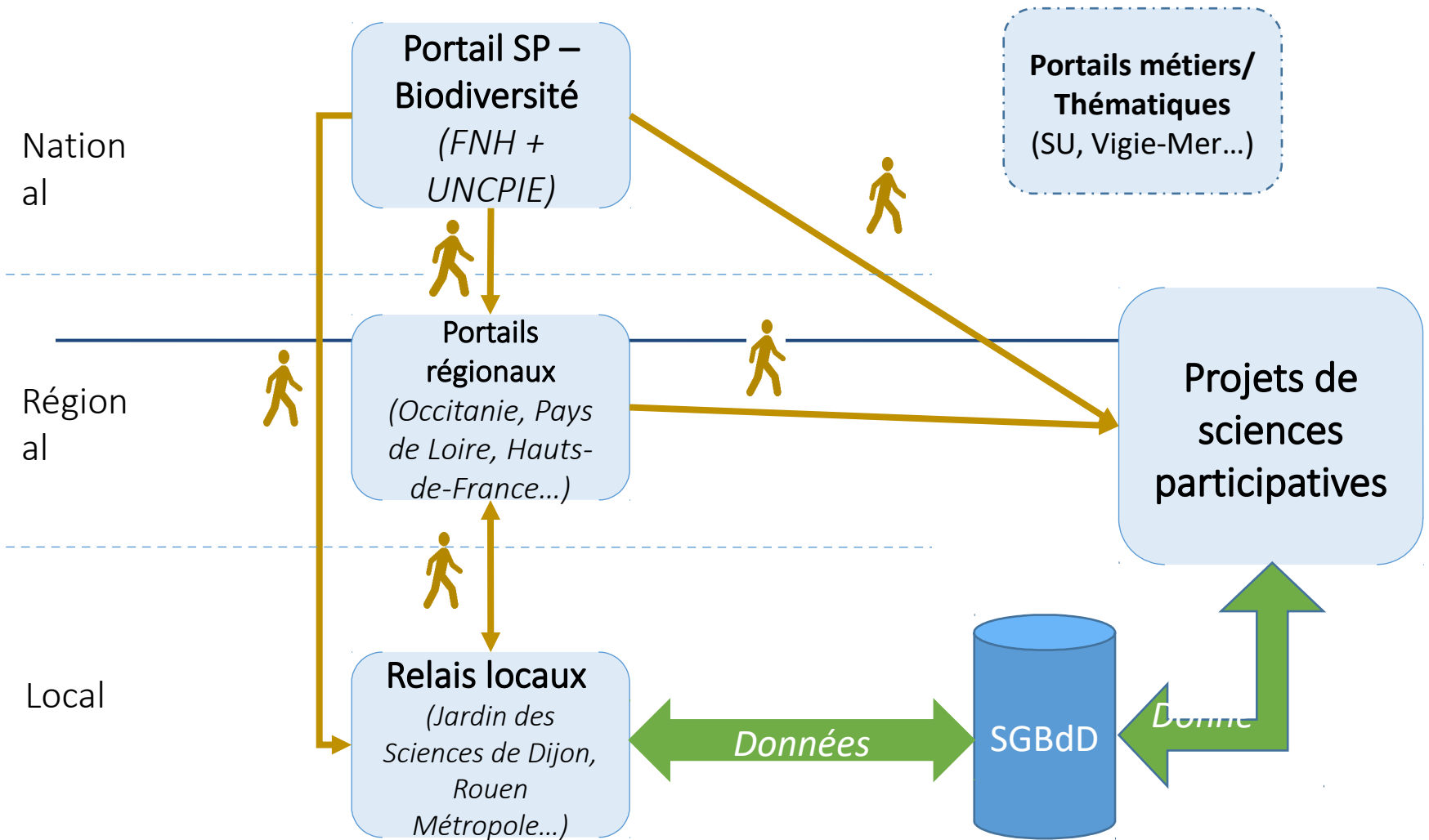
Le *Citizen Science Framework*

- **Un « *framework* » :**
 - Avec des fonctionnalités dédiées aux sciences participatives : participants, animateurs et porteurs de projet
 - Permettant de développer des sites de participation et d'animation à moindre coût
 - Pouvant bénéficier des services associés au SGBdD

- **Des services associés ?**
 - Circulation de l'information et des données
 - Mutualisation de tout ou partie de la maintenance et de l'hébergement

**3. Un constat :
Une animation des
programmes de sciences
participatives à plusieurs
échelles**

Structuration possible des SP-Biodiversité (à terme) (notre perception aujourd'hui)



« Portail » : bénéfices attendus

Pour les porteurs de projets :

- Création d'un réseau métier d'échange d'expérience et un centre de ressource pour de nouveaux projets
- Mutualisation d'une partie de l'animation

Pour l'ensemble des acteurs :

- Outil de visibilité et de communication vis-à-vis d'un public difficile à toucher par ailleurs

Pour les participants :

- Référencement des projets de SP
- Développement d'un sentiment d'appartenance à une méta-communauté

Différents projets en cours

- **Le Portail des sciences participatives en biodiversité**
(En cours de développement, animation conjointe UNCPIC et FNH)

The screenshot shows the homepage of the OPEN portal. At the top, there is a navigation bar with the logo 'OPEN' and several menu items: 'RECHERCHES PARTICIPATIVES', 'ÉQUIPES', 'ACTES', 'COMMUNAUTÉ', 'BLOG', and 'FAQ'. Below the navigation bar is a large banner image of purple flowers with the text 'JE RECHERCHE' in a circular graphic. A red bar at the bottom of the banner displays '18424 participants' and '253 observations'. Below the banner, a small text block reads: 'OPEN est le premier portail des sciences participatives au service de la biodiversité. Il permet à tous les citoyens d'identifier des enjeux d'observation et de participer, contribuant ainsi à faire avancer la science.' The main content area is divided into several sections: 'JE PARTICIPE À UN PROGRAMME' with filters for 'LIEUX' and 'THÈMES', a map of France showing project locations, and a list of 'MES RECHERCHES' including 'L'effet papillon', 'L'hiver de la taupe', 'La migration des oiseaux', 'Paris et les abeilles', and 'Les fleurs des champs'. Below this is a 'MES ACTUALITÉS' section with news items like 'Les établissements de l'INRA', 'La charte des sciences participatives est signée', 'Les observations de Vigie Nature', 'Jean Michel Ligez', and 'Les Bourgeois de Belgique'. To the right is a 'JE REJOINS LA COMMUNAUTÉ' section with icons for 'S'INSCRIRE', 'À QUOI PEUT-ON S'ATTENDRE?', and 'ET EN SUITE DE QUOI?' and a grid of project photos. At the bottom, there is a footer with social media icons and the text 'UNCPIC & FNH'.

Différents projets en cours

- **Le Portail des sciences participatives en biodiversité**
(En cours de développement, animation conjointe UNCPPIE et FNH)
- **Le Portail des sciences participatives liées à la recherche**
(Etude d'opportunité, sur le périmètre des établissements de Sorbonne Universités)
- **Le projet de Plateforme des acteurs de la Recherche culturelle participative**
(Financement Ministère de la Culture à partir du 1er janvier 2018, projet porté par Frédérique Chlous)



**Vers une infrastructure
numérique dédiée aux
sciences
participatives ?**

« 65 Millions d'Observateurs » : vers une infrastructure numérique nationale pour les Sciences Participatives



5 ans : 2015-2019
4,3 millions €
d'investissement

Une plateforme sociale :

Le *Citizen Science Framework* : une fabrique à sites web de participation (+ développements spécifiques)

1 an de recueil de besoin auprès de 100 acteurs des SP

Portails : mutualiser l'animation entre projets

Projets de sciences participatives

Relais locaux : distribuer l'animation

⇒ **Des outils-métiers pour animer les sciences participatives**

Une plateforme de données :

Le Système de Gestion de Bases de Données

- Catalogue de protocoles stabilisés
- Stockage et circulation des données
- Outils communs d'exploration des données

(Les données appartiennent aux porteurs de projet)

4. Phase 2 :

Comment étendre la participation

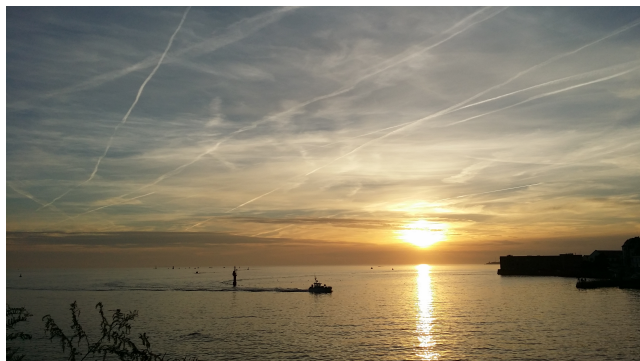
CONSULTATION ET ANALYSE DE DONNÉES DE PROJETS DE SCIENCES CITOYENNES

Yvan Le Bras Marie Delannoy Marianne Linares Grégoire Loïs Romain Julliard

CESCO UMR 7204

Projet 65 Millions d'observateurs

Station de biologie marine de Concarneau



65 millions d' observateurs analystes le projet « Galaxy-E »

Alan Amossé Valentin Chambon Eloïse Trigodet Thimothée Virgoulay Yvan Le Bras



CESCO UMR 7204 Projet 65 Millions d'observateurs
Station de biologie marine de Concarneau

« Datanami » en Sciences

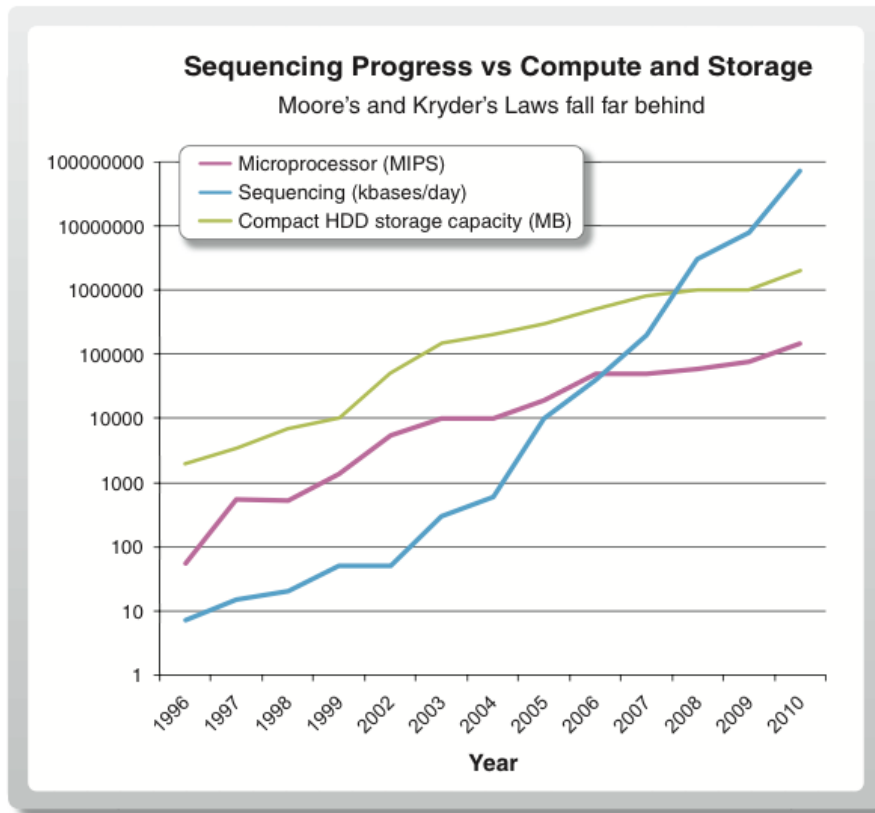


Fig. 1. A doubling of sequencing output every 9 months has outpaced and overtaken performance improvements within the disk storage and high-performance computation fields.

Kahn. On the future of genomic data. *Science* (2011) vol. 331(6018) pp. 728-9

- *omique : Séquençage nouvelle génération /
Protéomique /
Métabolomique
- Ecologie
- Donnée digitale
 - Quantité
 - Hétérogénéité
- Situation critique pour les laboratoires
- La solution = optimiser

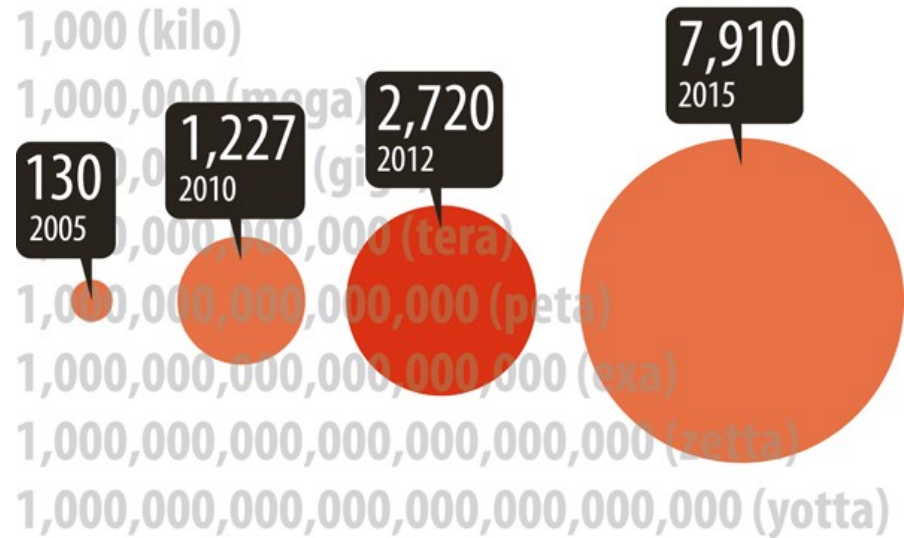
Perte de compétence analytique

“people with the skills to analyse data are scarce & will become scarcer”

Ludwig Siegele
Welcome to the yotta world
The economist, 2012

Exponential

Quantity of global digital data, exabytes

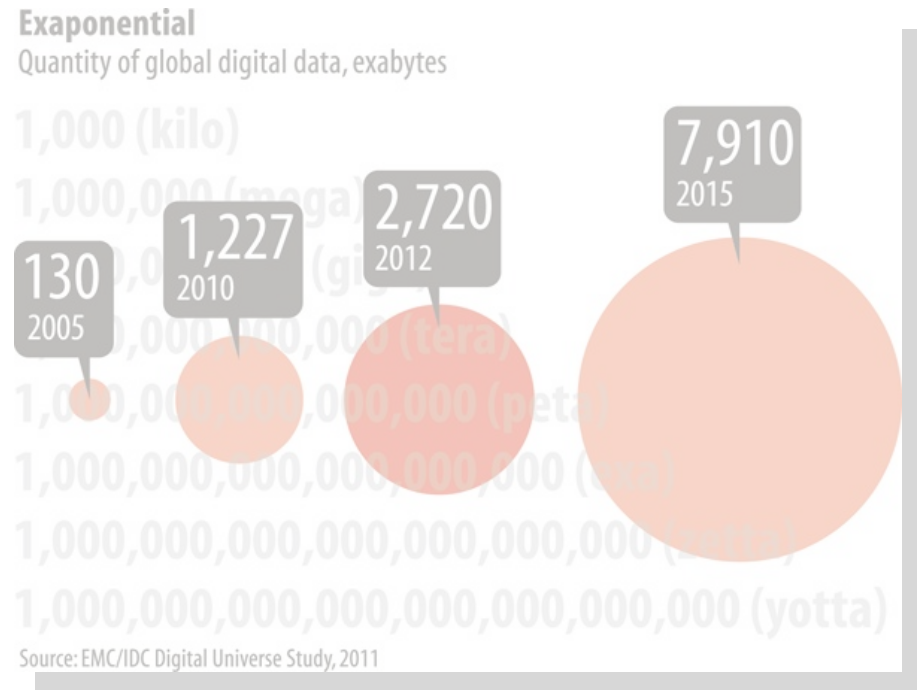


Source: EMC/IDC Digital Universe Study, 2011

Perte de compétence analytique

“people with the skills to analyse data are scarce & will become scarcer”

Ludwig Siegele
Welcome to the yotta world
The economist, 2012

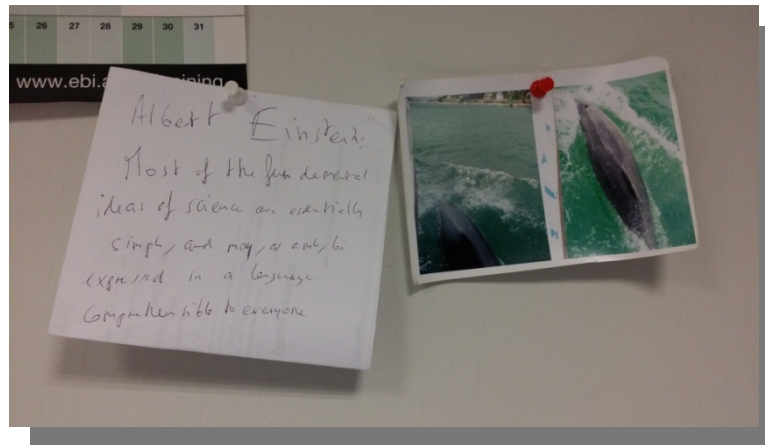


“One of the **grand challenges of data-intensive science is to facilitate knowledge discovery** by assisting humans and machines in their **discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows.**”

Perte de compétence analytique

- Faciliter l'analyse de données
 - Transfert de compétence (formation)
 - Accessibilité : outils plus utilisables
 - Reproductibilité : évaluation / ré-analyse / Portabilité
 - Transparence : Financements public/ citoyens & société

Perte de communication



« ...comprehensible to **everyone** »

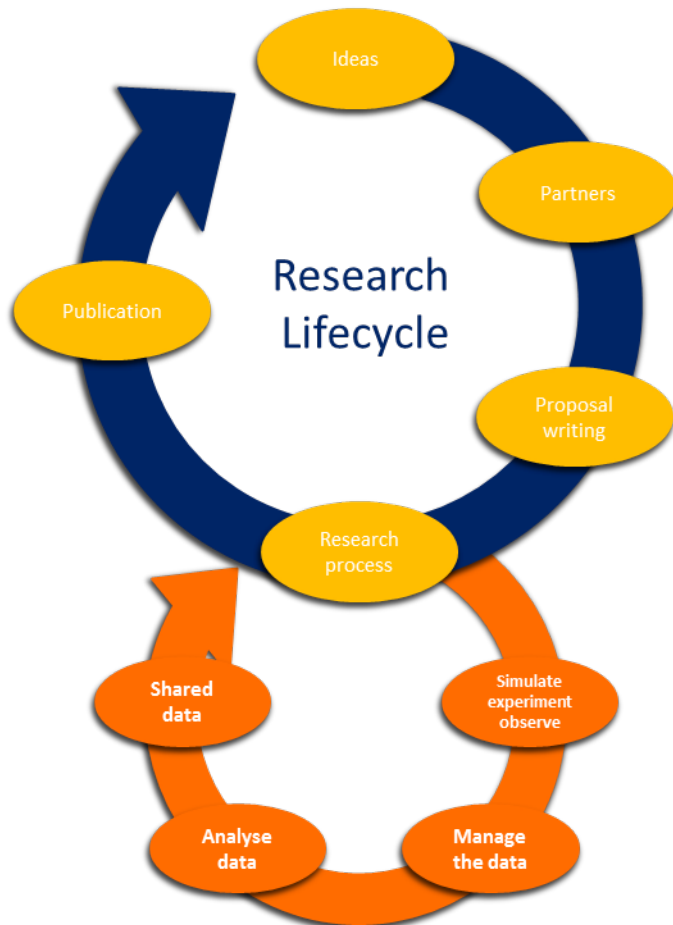
- Echanges

- Entre domaines scientifiques
- Entre scientifiques et acteurs du monde professionnel
 - Pêche
 - Agriculture
- Entre scientifiques et citoyens

Sciences participatives

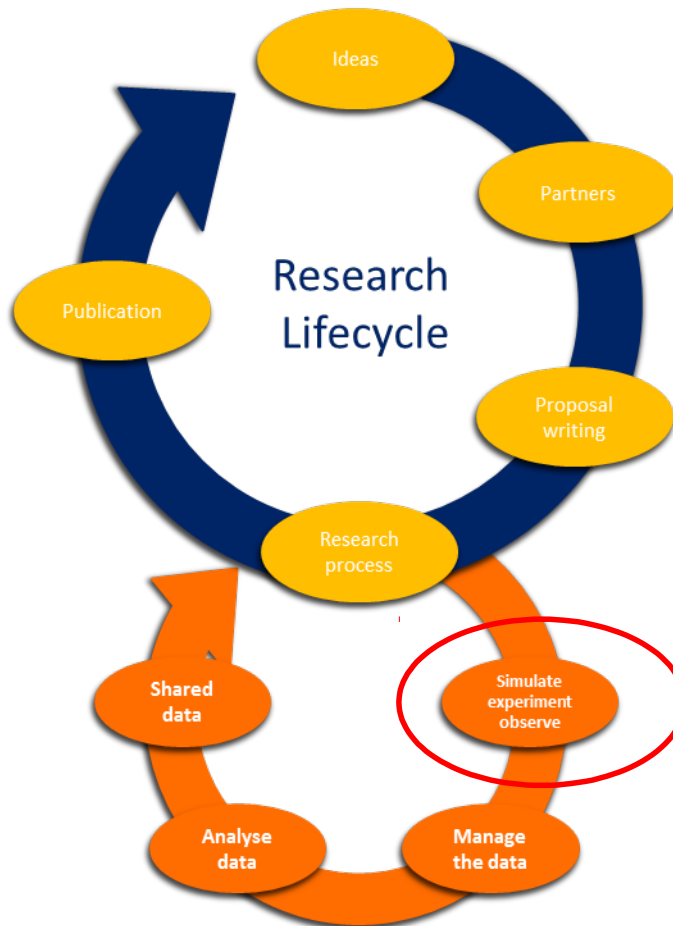
Contribution des citoyens dans les programmes de sciences participatives

Sciences participatives



Contribution des citoyens dans les programmes de sciences participatives

Sciences participatives

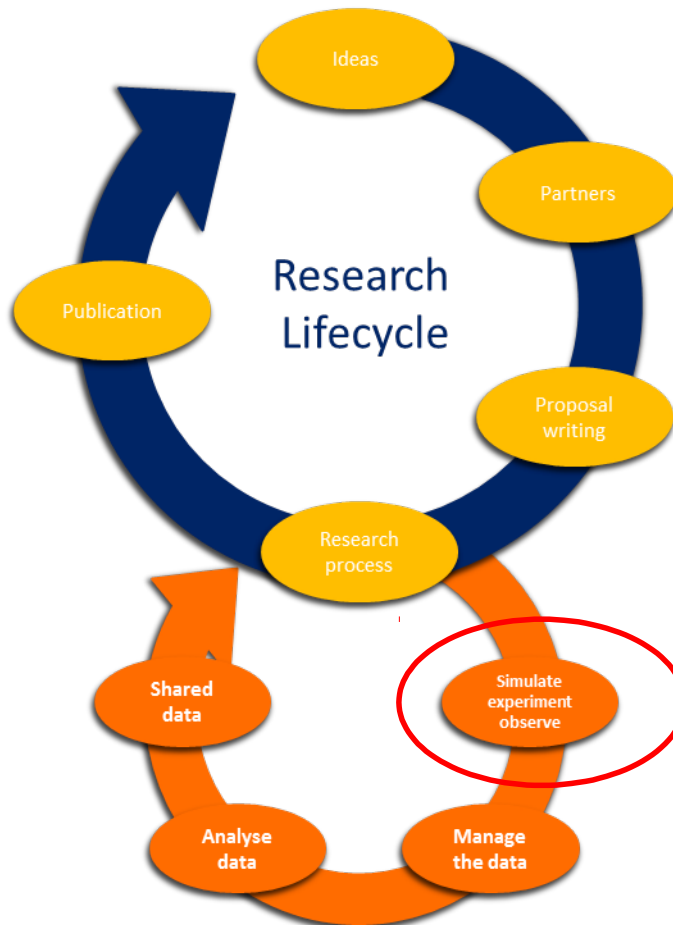


Contribution des citoyens dans les programmes de sciences participatives

Collecte de données

Cas général

Sciences participatives



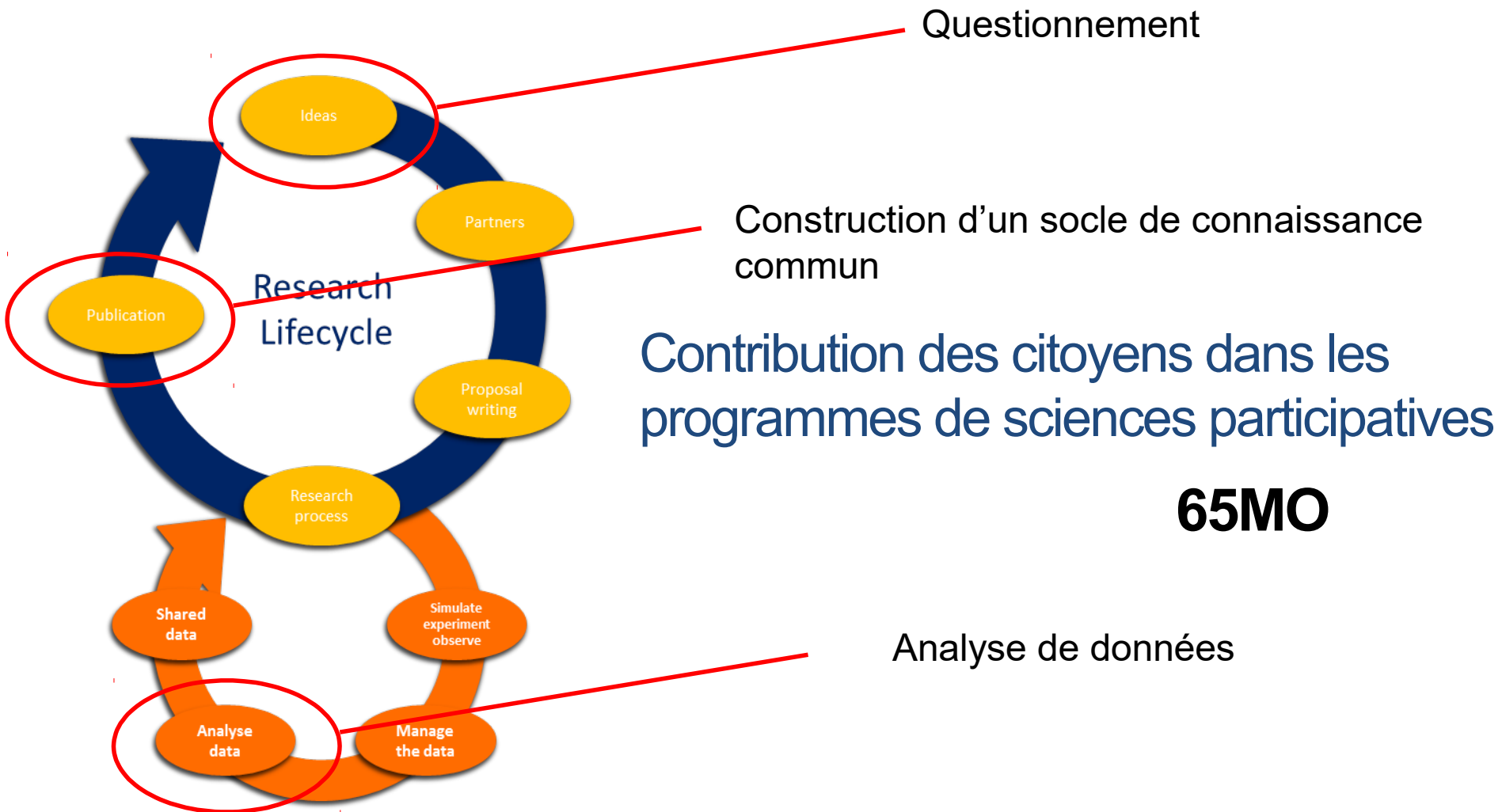
Contribution des citoyens dans les programmes de sciences participatives

Collecte de données

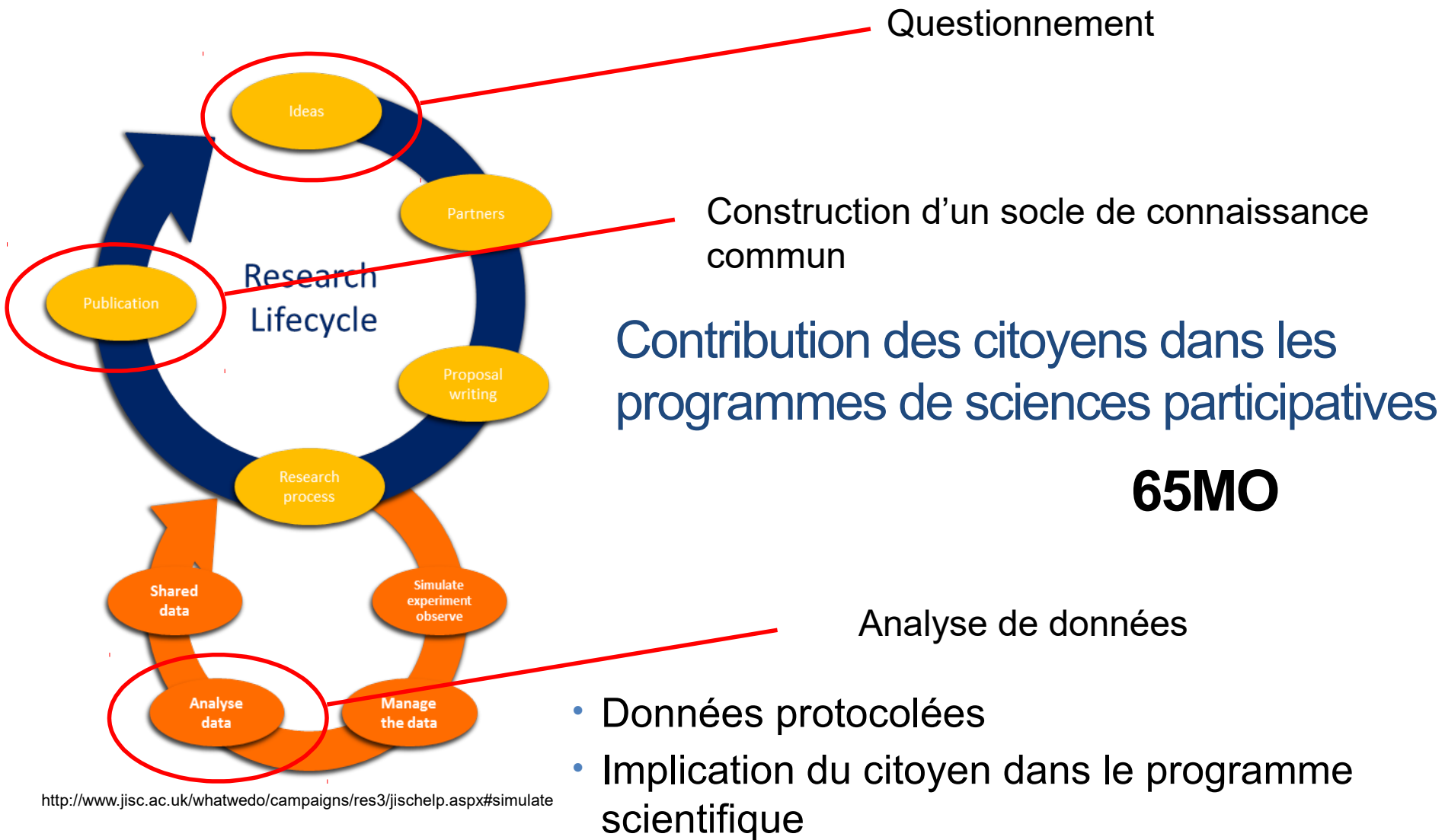
Cas général

- Nombre de producteurs de données important
- Main d'oeuvre « gratuite »

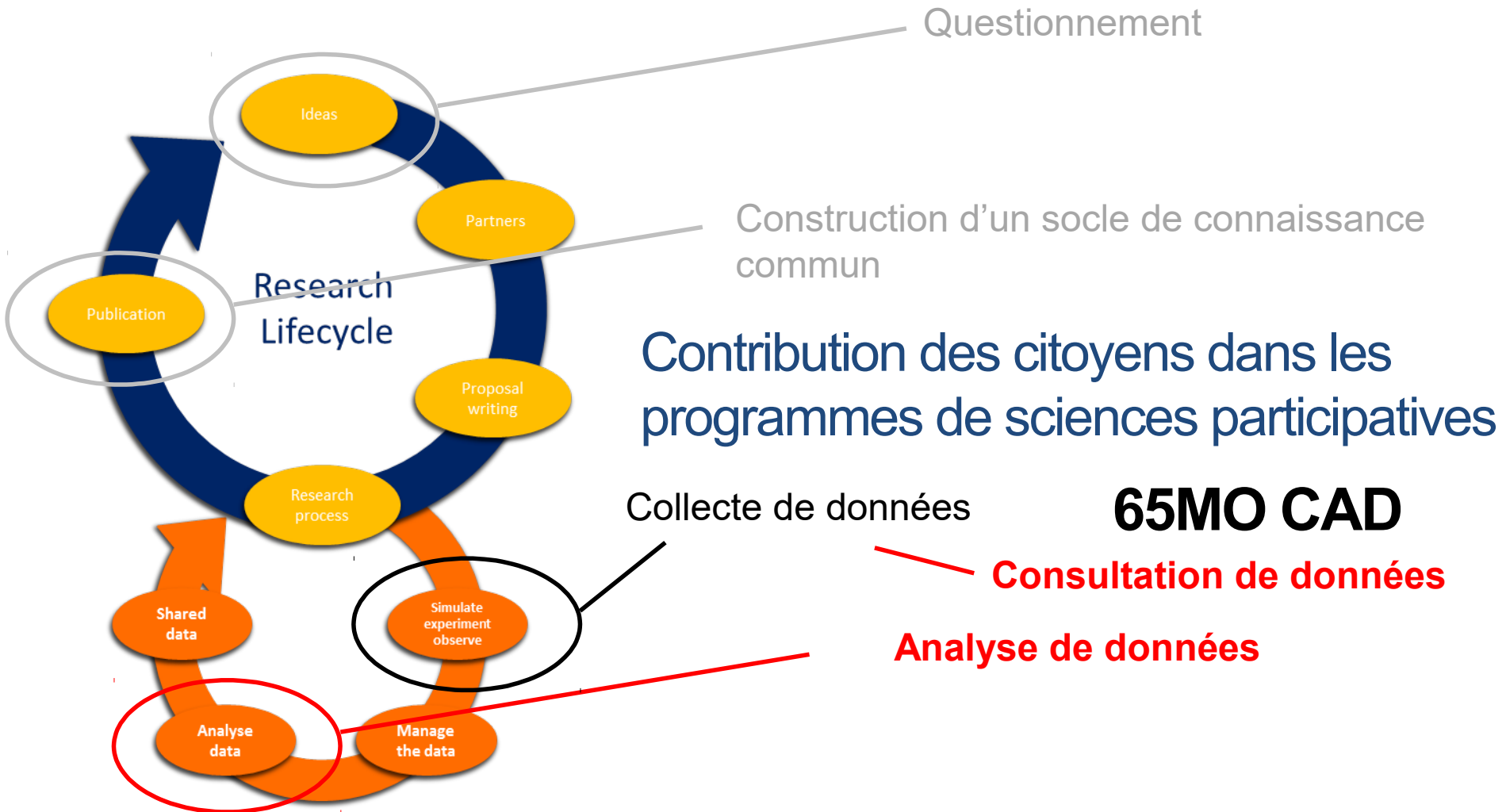
Sciences participatives



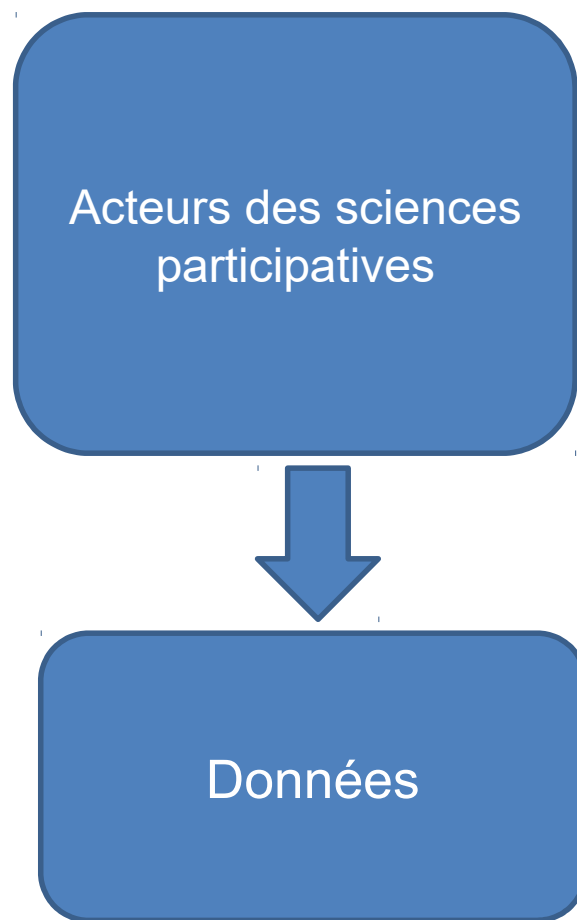
Sciences participatives



Sciences participatives



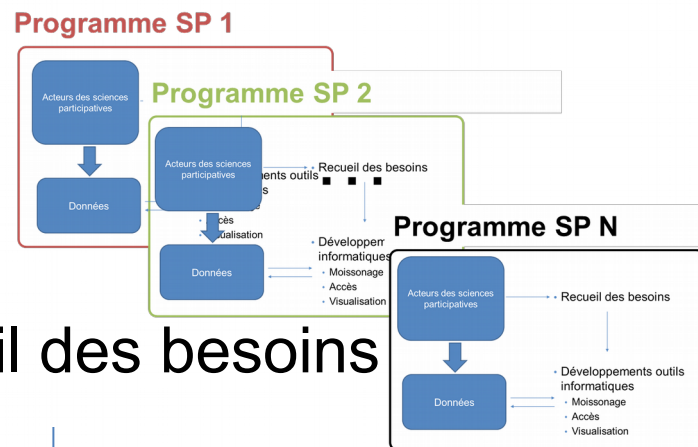
Objectifs du projet



• Recueil des besoins

• Développements outils informatiques

- Moissonnage
- Accès
- Visualisation
- Analyse



Enjeux

- Scientifiques

- Principes de données « FAIR »
 - Findable, Accessible, Interoperable, and Re-usable
- Passage à l'échelle
 - Faire face au datanami
- Nouvelles perspectives
 - Nouvelles idées
 - Nouveaux projets



- Sociétaux

- Améliorer la connaissance du monde environnant
 - les citoyens
 - les scientifiques
 - les professionnels usagers de l'environnement

- Environnementaux

- Meilleure connaissance ↔ meilleure gestion / conservation
- Meilleure utilisation des infrastructures

CONSULTATION & ANALYSE : SYNTHÈSE DU RECUEIL DE BESOINS

Classification au sein d'un « Workflow » d'analyse générique

1. Manipulation de bdd / Extraction de bdd
 - a) Interrogation / extraction simple de bdd existantes utilisable en un clic
2. Vérification / correction / standardisation des données par table
 - a) Outil de type [OpenRefine](#)
3. Manipulation de fichiers (un fichier = 1 table)
 - a) Modification encodage si besoin (ANSI / ISOXXXX vers UTF-8)
 - b) Modification formatage si besoin (csv, tsv, tabulé...)
 - c) Jointure de fichiers par ID
 - d) Sélection / Filtre/ Tri des données, récupération d'identifiants
 - e) Gestion des données manquantes / imputation de données
4. Statistiques descriptives & statistiques simples
 - a) Statistiques simples sur abondances
 - b) calculer d'indicateurs, d'indices de biodiversité
 - c) calcul de tendances temporelles
 - d) outils de traitement de signaux numériques (sons, images, vidéos) : classification, segmentation, reconnaissance automatique d'espèces, de patterns...
5. Restitution/Représentation graphique
 - a) Outils de prévisualisation des données, notamment pour fichiers csv/tsv extraits de bdd
 - b) Outils cartographiques pour données géoréférencées
6. Cartes de distribution
7. Cartes d'abondance
 - a) Outils graphiques (proche restitution et/ou mutualisé avec restitution) : histogrammes, nuages de points, boîtes à moustaches, ...
 - b) Tracking GPS en direct / récupération automatique des coordonnées GPS émises par un balise
8. Analyses statistiques « classiques »
 - a) Tests de corrélation
 - b) Représentation sous forme ACP / AFC (AUSSI section 5)
 - c) ANOVA
 - d) Croisement **dates** et/ou **données économiques** et/ou **données géographiques**
9. Modélisation notamment modélisation linéaire généralisée
 - a) Modélisation (observations en rapport avec l'attendu en fonction de l'environnement, modélisation de niche, ...)
 - b) Modélisation d'une phénologie en fonction des abondances par zones et période : ex variation été/hiver pour fauvette à tête noire
 - c) **Tendances temporelles** (espèces en expansion ou en déclin). Utilisé pour suivi de populations.
 - d) **Analyses spatiales** (impact occupation des sols)
 - e) **Influence** réchauffement climatique, météo, mode de gestion (fauche tardive), de la PAC sur oiseau des milieux agricoles,
10. Parcours / Interrogation des données via représentation sous forme de graphe
 - a) Utilisation d'outils d'interrogation de graphes (Gephi/Tulipe ou Askomics) (cf IndexMeed)

Extraction de données contenues dans des bases de données



Vérification / Correction / Standardisation des données



Manipulation / Formatage des jeux de données



Représentation graphique (nuages de points, histogrammes)



Représentation cartographique

Description des données et statistiques simples (min, max...)



Analyses statistiques (tests de corrélations, comparaisons, ...)



Analyses macro échelle par modélisation (calcul d'indices d'abondance, de tendances temporelles, ...)

1. Prétraitement des données

2. Restitution visuelle

3. Traitement analytique

Extraction de données contenues dans des bases de données



Vérification / Correction / Standardisation des données



Manipulation / Formatage des jeux de données



Description des données et statistiques simples (min, max...)



Analyses statistiques (tests de corrélations, comparaisons, ...)

1. Prétraitement des données

2. Restitution visuelle

3. Traitement analytique

4. Approche sémantique

Intégration des données sous forme de graphe



Interrogation intuitive du graphe de données

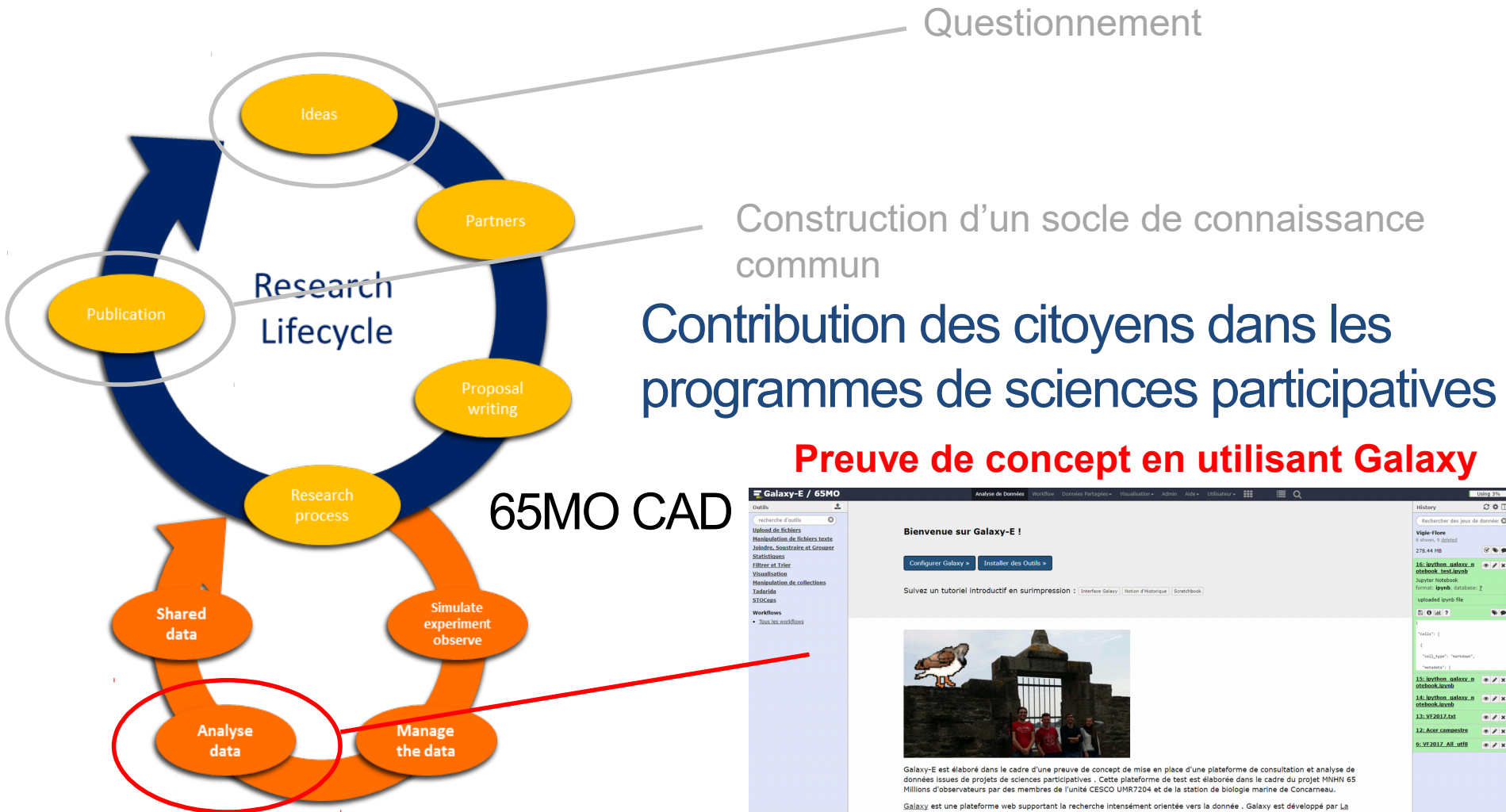
Représentation graphique (nuages de points, histogrammes)

Représentation cartographique

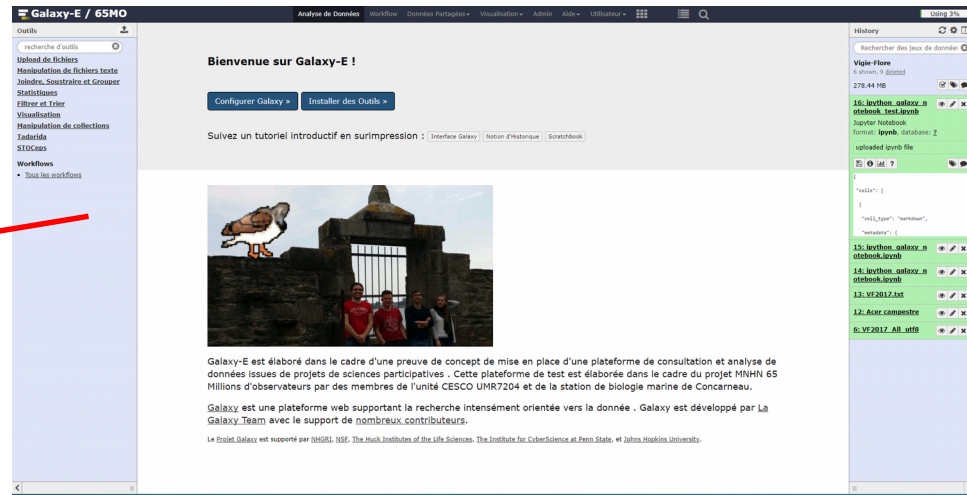
Analyses macro échelle par modélisation (calcul d'indices d'abondance, de tendances temporelles, ...)



Sciences participatives



<http://www.jisc.ac.uk/whatwedo/campaigns/res3/jiscshelp.aspx#simulate>



Galaxy-E, un Galaxy pour l'écologie

Galaxy-E / 65MO

Analyse de Données Workflow Données Partagées Visualisation Admin Aide Utilisateur

Outils

recherche d'outils

Upload de fichiers
Manipulation de fichiers texte
Joindre, Soustraire et Grouper
Statistiques
Filtrer et Trier
Visualisation
Manipulation de collections
Tadarida
STOCeps

Workflows

- Tous les workflows

Bienvenue sur Galaxy-E !

Configurer Galaxy » Installer des Outils »

Suivez un tutoriel introductif en surimpression : Interface Galaxy Notion d'Historique Scratchbook

Galaxy-E est élaboré dans le cadre d'une preuve de concept de mise en place d'une plateforme de consultation et analyse de données issues de projets de sciences participatives . Cette plateforme de test est élaborée dans le cadre du projet MNHN 65 Millions d'observateurs par des membres de l'unité CESCO UMR7204 et de la station de biologie marine de Concarneau.

Galaxy est une plateforme web supportant la recherche intensément orientée vers la donnée . Galaxy est développé par La Galaxy Team avec le support de nombreux contributeurs.

Le Projet Galaxy est supporté par NHGRI, NSF, The Huck Institutes of the Life Sciences, The Institute for CyberScience at Penn State, et Johns Hopkins University.

History

Rechercher des Jeux de données

Vigie-Flore
6 shown, 9 deleted
278.44 MB

16: ipython_galaxy_notebook_test.ipynb
Jupyter Notebook
format: ipynb, database: ?
uploaded ipynb file

```
{
  "cells": [
    {
      "cell_type": "markdown",
      "metadata": {
```

15: ipython_galaxy_notebook.ipynb
14: ipython_galaxy_notebook.ipynb
13: VF2017.txt
12: Acer campestre
6: VF2017 All utf8

Invocation d'un outil

Panneau d'outils

Historique des données-> Workflows

The screenshot displays the Galaxy-E / 65MC interface. The main panel shows the configuration for the 'Tadarida-d Toolbox for Animal Detection on Acoustic Recordings - Detection and Feature extraction part (Galaxy Version 1.0)'. The configuration includes a 'Sound file(s)' field with '1: 20160730_005625.wav', an 'Advanced Options' section with 'List of features' set to '1', 'Time Expansion Factor' set to '10', and 'Frequency Bands' set to '1'. A 'Citations' section is also visible, providing a reference to the tool's publication. The right-hand panel shows the 'History' section, listing imported data and workflow outputs, including '17: Trouver et Remplacer des patterns on data 3' and '3: output/20160730_005625', with details on line counts and formats.

Tadarida-d Toolbox for Animal Detection on Acoustic Recordings - Detection and Feature extraction part (Galaxy Version 1.0)

Sound file(s)
1: 20160730_005625.wav
Sound file(s) in Waveform Audio File Format (.wav)

Advanced Options

List of features
1

Sets the list of features to be extracted on each detected sound events. More informations availables at https://github.com/YvesBas/Tadarida-D/blob/master/Manual_Tadarida-D.odt (see appendix)

Time Expansion Factor
10
10 for 10-times expanded .wav files (most commonly used in bat monitoring) or 1 for direct recordings (only 10 or 1 possible).

Frequency Bands
1
Set the frequency bands to be used. 1 allows to treat high frequencies (8 to 250 kHz) whereas 2 treats low frequencies (0.8 to 25 kHz).

Execute

Citations Show BibTeX

Bas, Yves and Bas, Didier and Julien, Jean-François (2017). Tadarida: A Toolbox for Animal Detection on Acoustic Recordings. In *Journal of Open Research Software*, 5. [[doi:10.5334/jors.154](https://doi.org/10.5334/jors.154)][[Link](#)]

History

Rechercher des jeux de données

imported: Test Tadarida
4 shown, 12 deleted
26.48 MB

17: Trouver et Remplacer des patterns on data 3

3: output/20160730_005625
4 lines, 1 comments
format: csv, database: 2

randomForest 4.6-12
Type rfNews() to see new features/changes/bug fixes.

1.Group.1	2.Barbar	3.Eptser	4.Ins
"Group.1"	"Barbar"	"Eptser"	"Inss"
"dataset_533.dat"	0.092	0.446	0.406
"dataset_533.dat"	0.012	0.18	0.146
"dataset_533.dat"	0.074	0.462	0.026
"dataset_533.dat"	0	0.002	0.006

2: output/20160730_005625
1,255 lines
format: tabular, database: 2

Salut

1	2	3	4	5
Filename	CallNum	Version	fileDur	
dataset_533.dat	0	1	5.16	2
dataset_533.dat	1	1	5.16	2
dataset_533.dat	2	1	5.16	2
dataset_533.dat	3	1	5.16	2

1: 20160730_005625.wav

Formulaire de l'outil permettant de préciser les données d'entrées et les paramètres

Possibilité d'utiliser du code

Panneau d'outils

Historique des données -> Workflows

The screenshot displays the Galaxy-E / 65MO interface. The main area is a Jupyter notebook titled 'ipython_galaxy_notebook (autosaved)'. The notebook content includes a welcome message and two code cells:

```
In [3]: handle = open(get(3), 'r')
```

```
In [4]: descr=file("toto.txt", 'w')
i = 0
while i < 10:
    lines = handle.readline()
    descr.write(str(lines))
    i = i + 1
descr.close()
put ('toto.txt')
```

The right-hand side of the interface shows a 'History' panel with a search bar and a list of data objects. The objects listed include:

- imported: Test Tadarida (26.48 MB)
- 19: toto.txt (4 lines, 1 comment, format: csv, database: 2)
- 17: Trouver et Remplacer des patterns on data 3
- 3: output/2016073_0_005625 (4 lines, 1 comment, format: csv, database: 2)
- 2: output/2016073_0_005625 (1,255 lines, format: tabular, database: 2)

Blue arrows point from the text labels to the 'Outils' sidebar on the left and the 'History' panel on the right. A blue arrow also points from the bottom text to the code cells in the notebook.

Possibilité de coder dans plus de 30 langages dont R, Python, ... pour travailler sur les données

Vérification et correction de données

Panneau d'outils

Historique des données-> Workflows

Invocation d'OpenRefine
Un serveur est lancé pour traiter le jeu de données courant (ici « 3: Filter on data 2 » soit la 3^{ème} étape de l'historique)

Vérification et correction de données

Panneau d'outils

Historique des données-> Workflows

The screenshot displays the Galaxy-E / 65MO interface. The top navigation bar includes 'Analyse de Données', 'Workflow', 'Données Partagées', 'Visualisation', 'Admin', 'Aide', 'Utilisateur', and 'Using 3%'. The left sidebar contains a search bar for tools and a list of tool categories: Upload de fichiers, Manipulation de fichiers texte, Joindre, Soustraire et Grouper, Statistiques, Filtrer et Trier, Visualisation, Manipulation de collections, Tadarida, STOCeps, and Workflows. The main area shows a table of 2495 records with columns for DECADE, ID_OBS, JOUR, MOIS, ANNEE, HEURE, MATIN-AM, DUREE, ZONE GEOGRA, POS EXACTE L1, POS EXACTE L2, and POS EXACT. A 'Using facets and filters' tooltip is visible. On the right, the 'History' panel shows a list of workflows, including 'wf complet APECS sans regex' and several 'Count on data' and 'Filter on data' workflows. A workflow menu is open, showing options like 'Charts', 'Scatterplot', 'AskOmics', 'Ethercalc', 'Jupyter', 'openrefine', and 'RStudio'.

Utilisation d'OpenRefine pour traiter les données puis les réexporter dans l'historique courant

Vérification et correction de données

Panneau d'outils

Historique des données-> Workflows

The screenshot displays the Galaxy-E / 65MO interface. On the left, a sidebar contains a search bar and various tool categories: Upload de fichiers, Manipulation de fichiers texte, Joindre, Soustraire et Grouper, Statistiques, Filtrer et Trier, Visualisation, Manipulation de collections, Tadarida, STOCeps, and Workflows. The main area shows a data table with 2495 records, filtered by 'Facet / Filter'. A tooltip 'Using facets and filters' explains how to use the filters. An 'Export project' menu is open, showing options like Tab-separated value, Comma-separated value, HTML table, Excel (.xls), Excel 2007+ (.xlsx), ODF spreadsheet, Triple loader, MQLWrite, Custom tabular exporter..., and Templating... The right sidebar shows a 'History' panel with a list of workflows, including 'wf complet APECS sans regex' and several 'Filter on data' workflows. A text box at the bottom center reads: 'Utilisation d'OpenRefine pour traiter les données puis les réexporter dans l'historique courant'.

Galaxy-E / 65MO

recherche d'outils

Upload de fichiers

Manipulation de fichiers texte

Joindre, Soustraire et Grouper

Statistiques

Filtrer et Trier

Visualisation

Manipulation de collections

Tadarida

STOCeps

Workflows

- Tous les workflows

OPEN Refine Filter Permalink

Facet / Filter Undo / Redo

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts

2495 records

Show as: rows records

	All	DECADE	ID_OBS	JOUR	MOIS	ANNEE	HEURE	MATIN-AM	DUREE	ZONE GEOGRA									
1.	1		19	6	1	2002	14.00.00	Après-Midi	15	Manche									
2.	1		20	10	1	2002				Atlantique									
3.	1		21	2	3	2002	10.00.00	Matin		Méditerranée									
4.	3		22	21	4	2002			20	Atlantique	47	46	176						
5.			23	24	4	2002	11.00.00	Matin	5	Méditerranée									
6.	3		24	26	4	2002	10.00.00	Matin	30	Atlantique	46	41							
7.	1		26	1	5	2002	16.00.00	Après-Midi	15	Atlantique									
8.	1		29	6	5	2002	10.30.00	Matin		Atlantique									
9.	1		30	10	5	2002	14.00.00	Après-Midi		Atlantique									
10.	1		31	10	5	2002	08.50.00	Matin	2	Atlantique	47	40	81						

Export project

- Tab-separated value
- Comma-separated value
- HTML table
- Excel (.xls)
- Excel 2007+ (.xlsx)
- ODF spreadsheet
- Triple loader
- MQLWrite
- Custom tabular exporter...
- Templating...

History

Rechercher des jeux de données:

wf complet APECS sans regex

27 shown, 31 deleted

12.1 MB

- 10: Count on data 5
- 9: Count on data 4
- 8: Count on data 4
- 7: Sort on data 4
- 6: Filter on data 3
- 5: Filter on data 3
- 4: Filter on data 3
- 3: Filter on data 2

>1997

2,523 lines

format: tabular, database: ?

Filtering with c5> 1997, kept 85.55% of 2949 valid lines (2949 total lines).

- 1: Join two Datasets on data 3 and data 2
- 2: Join two Datasets on data 1

Charts

- Scatterplot
- AskOmic
- Ethercalc
- Jupyter
- openrefine
- RStudio

Utilisation d'OpenRefine pour traiter les données puis les réexporter dans l'historique courant

Vérification et correction de données

Panneau d'outils

Données exportées dans un nouveau jeu de données peuplant l'historique

The screenshot displays the Galaxy-E / 65MC interface. At the top, there's a navigation bar with 'Analyse de Données', 'Workflow', 'Données Partagées', 'Visualisation', 'Admin', 'Aide', 'Utilisateur', and a search icon. Below this is a warning banner: 'This dataset is large and only the first megabyte is shown below. Show all | Save'. On the left, a sidebar contains tool categories: 'Outils', 'Upload de fichiers', 'Manipulation de fichiers texte', 'Statistiques', 'Filtrer et Trier', 'Visualisation', 'Manipulation de collections', 'Tadarida', 'STOCeps', 'Workflows', and 'Tous les workflows'. The main area shows a data table with columns: DECADE, ID_OBS, JOUR, MOIS, ANNEE, HEURE, MATIN-AM, DUREE, ZONE GEOGRAPHIQUE, POS EXACTE LAT DEG, POS EXACTE LAT MIN, POS EXACTE LAT SEC, POS EXACTE LONG DEG, POS EXACTE LONG MIN, POS EXACTE LONG SEC, POS EXACTE LONG DEG, POS EXACTE LONG MIN, POS EXACTE LONG SEC. The table contains multiple rows of data. On the right, a 'History' panel shows a list of workflows, including 'wf complet APECS sans regex', '59: Filter refine', '54: Tadarida classifi...', '53: fichier export...', '52: fichier export...', '51: fichier to get...', '23: Pasted Entry', '21: Cut on data 19', '20: Cut on data 18', '19: Sort on data 17', '18: Join two Datas...', and '17: Join two Datas...'. A callout box points to the '59: Filter refine' workflow, indicating that data is exported into a new dataset.

DECADE	ID_OBS	JOUR	MOIS	ANNEE	HEURE	MATIN-AM	DUREE	ZONE GEOGRAPHIQUE	POS EXACTE LAT DEG	POS EXACTE LAT MIN	POS EXACTE LAT SEC	POS EXACTE LONG DEG	POS EXACTE LONG MIN	POS EXACTE LONG SEC	POS EXACTE LONG DEG	POS EXACTE LONG MIN	POS EXACTE LONG SEC							
1	19	6	1	2002	14:00:00	Apr	15	Manche	48	38	500	2	47	500	false	48	35	48.5833333333333	2	50	-2.83333333333333	54		
1	20	10	1	2002		Atlantique	44	34	674	1	14	584	false	44	35	44.5833333333333	1	10	-1.16666666666667	La Teste	54			
1	21	2	3	2002	10:00:00	Matin	15	Méditerranée																
3	22	21	4	2002		20	Atlantique	47	46	176	3	41	95	false	47	45	47.75	3	40	-3.66666666666667	Mo	lan Sur Mer	54	
3	23	24	4	2002	11:00:00	Matin	5	Méditerranée																
3	24	26	4	2002	10:00:00	Matin	30	Atlantique	46	41		2	19	false	46	45	46.75	2	20	-2.33333333333333	Ile D'yeu	54		
1	26	1	5	2002	16:00:00	Apr	15	Atlantique																
1	29	6	5	2002	10:30:00	Matin		Atlantique																
1	30	10	5	2002	14:00:00	Apr		Atlantique																
1	31	10	5	2002	08:50:00	Matin	2	Atlantique	47	40	81	3	31	72	false	47	45	47.75	3	30	-3.5	Ile De Groix	Mc	54
1	32	10	5	2002	10:40:00	Matin	15	Atlantique	47	39		3	28	false	47	35	47.5833333333333	3	30	-4.16666666666667	Ile De Groix	Mc	54	
2	33	11	5	2002	16:30:00	Apr	2	Mer d'Iroise																
2	34	16	5	2002	09:20:00	Matin	10	Atlantique																
2	35	18	5	2002	15:00:00	Apr	5	Atlantique																
2	36	20	5	2002		3	Atlantique	47	23		3	5	false	47	25	47.4166666666667	3	0	-3	Belle Ile	Mc	54		
3	37	26	5	2002		5	Atlantique	47	28		3	15	false	47	25	47.4166666666667	3	10	-3.16666666666667	Quiberon	Lc	54		
3	38	29	5	2002	17:30:00	Apr	2	Atlantique																
3	39	31	5	2002	11:00:00	Matin	30	Atlantique																
3	40	31	5	2002	10:00:00	Matin	20	Atlantique																
1	41	1	6	2002	15:00:00	Apr	20	Atlantique	46	36		2	4	false	46	46	35	46.5833333333333	2	0	-2			54
1	41	1	6	2002	15:00:00	Apr	20	Atlantique	46	36		2	4	false	46	35	46.5833333333333	2	0	-2			54	
1	42	2	6	2002	11:30:00	Matin	30	Atlantique																
1	43	2	6	2002		Atlantique	47	20	663	3	6	536	false	47	25	47.4166666666667	3	10	-3.16666666666667	Belle Ile		54		
1	44	4	6	2002	07:00:00	Matin	10	Atlantique	47	20	134	3	8	506	false	47	25	47.4166666666667	3	10	-3.16666666666667	Belle Ile		54
1	45	6	6	2002	08:15:00	Matin	45	Atlantique																
1	46	8	6	2002	11:00:00	Matin		Mer d'Iroise																
2	47	12	7	2002	12:00:00	Apr		Mer d'Iroise																
2	48	19	7	2002	16:00:00	Apr		Manche	48	58		3	34	false	48	55	48.9166666666667	3	30	-3.5	Perros Guj		54	
3	49	22	7	2002	19:14:00	Apr		Atlantique	47	45		3	40	false	47	45	47.75	3	40	-3.66666666666667	Mo		54	
3	50	22	7	2002	15:00:00	Apr		Atlantique																
3	51	31	7	2002	17:10:00	Apr		Manche	49	0		4	1	false	48	55	48.9166666666667	4	0	-4	Ile De Bat		54	
						Matin		Atlantique																
						Apr		Mer d'Iroise	48	20	1	4	52	50	false	48	25	48.4166666666667	4	50	-4.83333333333333	Le		54
						Matin		Manche																
						Matin		Manche																
						Matin		Méditerranée																
						Apr		Manche	48	44	540	4	8	560	false	48	45	48.75	4	10	-4.16666666666667	Ile De Batz	Finist	54
						Matin		Atlantique																
						Apr		Atlantique	47	33		3	12	false	47	35	47.5833333333333	3	10	-3.16666666666667	Mo		54	
						Matin		Manche	48	53		3	59	false	48	55	48.9166666666667	4	0	-4	Ile De Batz	Fi	54	
						Apr		Atlantique	47	47		4	19	false	47	45	47.75	4	20	-4.33333333333333	Pe		54	
						Apr		Manche	48	42	474	4	12	732	false	48	45	48.75	4	10	-4.16666666666667	Plovecat		54
						Matin		Manche																
						Matin		Manche																
						Apr		Atlantique	46	5	720	1	10	250	false	46	5	46.0833333333333	1	10	-1.16666666666667			54
						Apr		Atlantique																
						Apr		Méditerranée	42	49	500	-5	-33	-800	false	42	45	42.75	-5	-30	-5.5			54
						Apr		Méditerranée	42	44	100	-5	-38	-200	false	42	45	42.75	-5	-40	-5.66666666666667			54
						Atlantique																		

Visualisation du jeu de données exporté

Manipulation de fichiers

Panneau d'outils

Ensemble des étapes de filtre / tri des données

Galaxy-E / 65MO

Analyse de Données Workflow Données Partagées Visualisation Admin Aide Utilisateur Using 3%

Outils

recherche d'outils

Upload de fichiers

Manipulation de fichiers texte

Joindre, Soustraire et Grouper

Statistiques

Filtrer et Trier

Visualisation

Manipulation de collections

Tadarida

STOCeps

Workflows

- Tous les workflows

1	2	3	4
MOIS	1	1	1
10	11	2	3
11	12	1	6
12	12	3	1
1	20	2	4
3	79	3	7
7	322	2	2
4	336	6	12
6	512	2	16
5	659	8	39

History

Rechercher des jeux de données:

Workflow constructed from history 'Projet de stage Éloïse APECS'

20 shown, 24 deleted

5.72 MB

44: Sélectionner les dernières lignes on data 17

10 lignes

format: tabular, database: 2

1	2	3	4
MOIS	1	1	1
10	11	2	3
11	12	1	6
12	12	3	1
1	20	2	4

40: Sélectionner les dernières lignes on data 2

18: Couper on data 16

Mois

17: Couper on data 15

Année

16: Joindre 2 jeux de données on data 1 2 and data 14

Mois

15: Joindre 2 jeux de données on data 1 1 and data 13

Année

14: Joindre 2 jeux de données on data 1 0 and data 8

Echouage Observation

13: Joindre 2 jeux d

Visualisation du jeu de données final filtré

Statistiques descriptives et simples

Panneau d'outils

The screenshot displays the Galaxy-E / 65MC web interface. The main panel shows the configuration for the 'STOCeps' tool, which is used for creating annual variation files and global trend files for species. The tool is set to use 'Données STOC' with the input '21: Année, abondance relative, ic_inferieur, ic_supérieur'. The interface includes a sidebar with navigation options like 'Outils', 'Upload de fichiers', and 'Workflows'. A 'History' panel on the right lists 14 generated datasets, including '21: Année, abondance relative, ic_inferieur, ic_supérieur' and '18: tendanceGlobalEspèceSTOCeps'. A blue arrow points from the 'Panneau d'outils' text to the tool configuration area.

STOCeps Création des fichiers des variations annuelles par espèces et de la tendance globale par espèce STOCeps. (Galaxy Version 1.0)

Données STOC

21: Année, abondance relative, ic_inferieur, ic_supérieur
Tableau issu des observations STOCeps

Execute

L'outil permet de générer deux jeux de données :

le fichier des variations annuelles par espèce :

- Ce fichier contient les sorties et ses interprétations du modèle statistique (glm variations d'abondances en fonction du carré et de l'année) qui permet de voir les variations inter-annuelles d'abondance des populations d'oiseaux pour chaque espèce. Ces grâce à ses valeurs que les représentations graphiques sont réalisés.

le fichier de la tendance globale par espèce :

- Ce fichier contient les sorties et ses interprétations du modèle statistique qui permet d'avoir la tendance globale de la variation d'abondance sur l'ensemble de la période de temps de l'espèce considérée. Le modèle utilisé est le même que pour les variations inter-annuelles, mais utilise les années comme une variable continue.

Exemple pour l'analyse des tendances globales et des variations annuelles d'abondance par espèces pour le dispositif STOCeps

Statistiques descriptives et simples

Fichiers de résultats

Galaxy-E / 65M0

Analyse de Données Workflow Données Partagées Visualisation Admin Aide Utilisateur Using 3%

imported: STOCeps: tendanceGlobalEspeceSTOCeps

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
id	code_espece	nom_espece	indicateur	nombre_annees	premiere_annee	derniere_annee	tendance	IC_inferieur	IC_superieur	pourcentage_variation	erreur_standard	p_value	significatif	categorie_
20170809-09H40	ACCNIS	Epervier d'Europe	non	13	2001	2014	1.021	0.959	1.089	30.806	0.0317	0.507	FALSE	Incertain
20170809-09H40	ACRSCI	Rousserolle effarvate	non	13	2001	2014	0.983	0.953	1.025	-19.63	0.0187	0.377	FALSE	Stable
20170809-09H40	AEGCAU	Mésange à longue queue	non	13	2001	2014	1.041	0.992	1.086	69.206	0.024	0.08	FALSE	Incertain
20170809-09H40	ALAARV	Alouette des champs	milieux agricoles	13	2001	2014	0.976	0.965	0.993	-26.71	0.0081	0.004	TRUE	Declin mod
20170809-09H40	ANAPLA	Canard colvert	non	13	2001	2014	1.054	1.02	1.083	97.781	0.0196	0.005	TRUE	Augmentat
20170809-09H40	ANTRRI	Pipit des arbres	non	13	2001	2014	0.996	0.977	1.019	-5.54	0.012	0.715	FALSE	Stable
20170809-09H40	APUAPU	Martinet noir	milieux batis	13	2001	2014	0.95	0.921	0.982	-48.396	0.0161	0.003	TRUE	Declin mod
20170809-09H40	ARDCIN	Héron cendré	non	13	2001	2014	0.985	0.95	1.026	-18.168	0.019	0.426	FALSE	Stable
20170809-09H40	BUTBUT	Buse variable	milieux agricoles	13	2001	2014	1.037	0.999	1.072	60.334	0.0186	0.044	TRUE	Augmentat
20170809-09H40	CARCAN	Linotte mélodieuse	milieux agricoles	13	2001	2014	0.837	0.805	0.866	-90.102	0.0188	0	TRUE	Fort declin
20170809-09H40	CARCAR	Chardonneret élégant	milieux batis	13	2001	2014	0.888	0.855	0.915	-78.696	0.0159	0	TRUE	Fort declin
20170809-09H40	CARCHL	Verdier d'Europe	milieux batis	13	2001	2014	0.904	0.885	0.918	-73.232	0.0082	0	TRUE	Fort declin
20170809-09H40	CERBRA	Grimpereau des jardins	milieux forestiers	13	2001	2014	0.974	0.954	1	-28.558	0.0127	0.048	TRUE	Declin mod
20170809-09H40	CETCET	Bouscarle de Cetti	non	13	2001	2014	0.956	0.922	0.99	-44.206	0.0191	0.025	TRUE	Declin mod
20170809-09H40	CIRAER	Busard des roseaux	non	13	2001	2014	0.821	0.783	0.855	-92.311	0.0249	0	TRUE	Fort declin
20170809-09H40	COLLIV	Pigeon biset	non	13	2001	2014	1.014	0.967	1.061	19.849	0.0275	0.607	FALSE	Incertain
20170809-09H40	COLPAL	Pigeon ramier	generaliste	13	2001	2014	1.033	1.022	1.042	53.411	0.0049	0	TRUE	Augmentat
20170809-09H40	CORCOR	Cornelle noire	generaliste	13	2001	2014	0.976	0.965	0.986	-26.777	0.0072	0.001	TRUE	Declin mod
20170809-09H40	CORMON	Choucas des tours	milieux batis	13	2001	2014	1.053	1.011	1.095	96.23	0.0202	0.007	TRUE	Augmentat
20170809-09H40	CUCCAN	Coucou gris	generaliste	13	2001	2014	1.022	1	1.039	33.332	0.01	0.024	TRUE	Augmentat

imported: STOCeps: variationsAnnuellesSTOCeps

1	2	3	4	5	6	7	8	9	10	11	12	13	14
id	code_espece	nom_espece	indicateur	annee	abondance_relative	IC_inferieur	IC_superieur	erreur_standard	p_value	significatif	nb_carre	nb_carre_presence	abondance
20170809-09H40	ACCNIS	Epervier d'Europe	non	2001	1	1	1	0	1	FALSE	15	4	4
20170809-09H40	ACCNIS	Epervier d'Europe	non	2002	0.216	0.092	0.694	0.547	0.005	TRUE	46	4	4
20170809-09H40	ACCNIS	Epervier d'Europe	non	2003	0.464	0.225	1.023	0.0989	0.094	FALSE	55	6	7
20170809-09H40	ACCNIS	Epervier d'Europe	non	2004	0.831	0.395	1.791	0.1924	0.655	FALSE	57	9	11
20170809-09H40	ACCNIS	Epervier d'Europe	non	2005	0.343	0.154	1.26	0.435	0.042	TRUE	55	4	4
20170809-09H40	ACCNIS	Epervier d'Europe	non	2006	0.658	0.271	1.374	0.1508	0.341	FALSE	56	7	7
20170809-09H40	ACCNIS	Epervier d'Europe	non	2007	0.278	0.093	0.67	0.3479	0.016	TRUE	52	3	3
20170809-09H40	ACCNIS	Epervier d'Europe	non	2008	0.157	0.03	0.623	0.2101	0.015	TRUE	44	1	1
20170809-09H40	ACCNIS	Epervier d'Europe	non	2009	0.416	0.189	1.302	0.0777	0.077	FALSE	48	4	4
20170809-09H40	ACCNIS	Epervier d'Europe	non	2010	0.859	0.396	2.367	0.182	0.728	FALSE	43	4	8
20170809-09H40	ACCNIS	Epervier d'Europe	non	2011	0.389	0.164	1.007	0.5147	0.116	FALSE	27	2	2
20170809-09H40	ACCNIS	Epervier d'Europe	non	2012	1.646	0.477	6.856	0.2533	0.444	FALSE	23	3	3
20170809-09H40	ACCNIS	Epervier d'Europe	non	2013	0.504	0.161	1.549	0.9958	0.258	FALSE	30	3	3
20170809-09H40	ACCNIS	Epervier d'Europe	non	2014	5.277	0.883	31.522	0.5052	0.098	FALSE	7	1	1

Représentation graphique simple

Panneau d'outils

Historique des données -> Workflows

The screenshot displays the Galaxy-E / 65MO interface. The main area is titled 'New Chart' and contains a form for creating a visualization. The 'Provide a title:' field is set to 'New Chart'. Below this, there are tabs for 'Suggested visualizations' and 'List of available visualizations'. The 'Suggested visualizations' tab is active, showing a grid of 15 visualization options: Bar diagram (NVD3), Horizontal (NVD3), Stacked horizontal (...), Stacked (NVD3), QLine chart (NVD3), QLine with focus (NVD...), QScatter plot (NVD3), QStacked area (NVD3), QExpanded (NVD3), QStream (NVD3), Pie chart (NVD3), Histogram (NVD3), QBox plot (jqPlot), QDiscrete Histogram (...), QHeatmap (Custom), QClustered Heatmap (C...), and Venn Diagram (benfre...).

On the left side, there is a 'Panneau d'outils' (Tools panel) with a search bar and several tool categories: 'Upload de fichiers', 'Manipulation de fichiers texte', 'Statistiques', 'Filtrer et Trier', 'Visualisation', 'Manipulation de collections', 'Tadarida', and 'STOCeps'. The 'Tadarida' category includes 'Tadarida-d' and 'Tadarida-c'. The 'STOCeps' category includes 'STOCeps2' and 'STOCeps'.

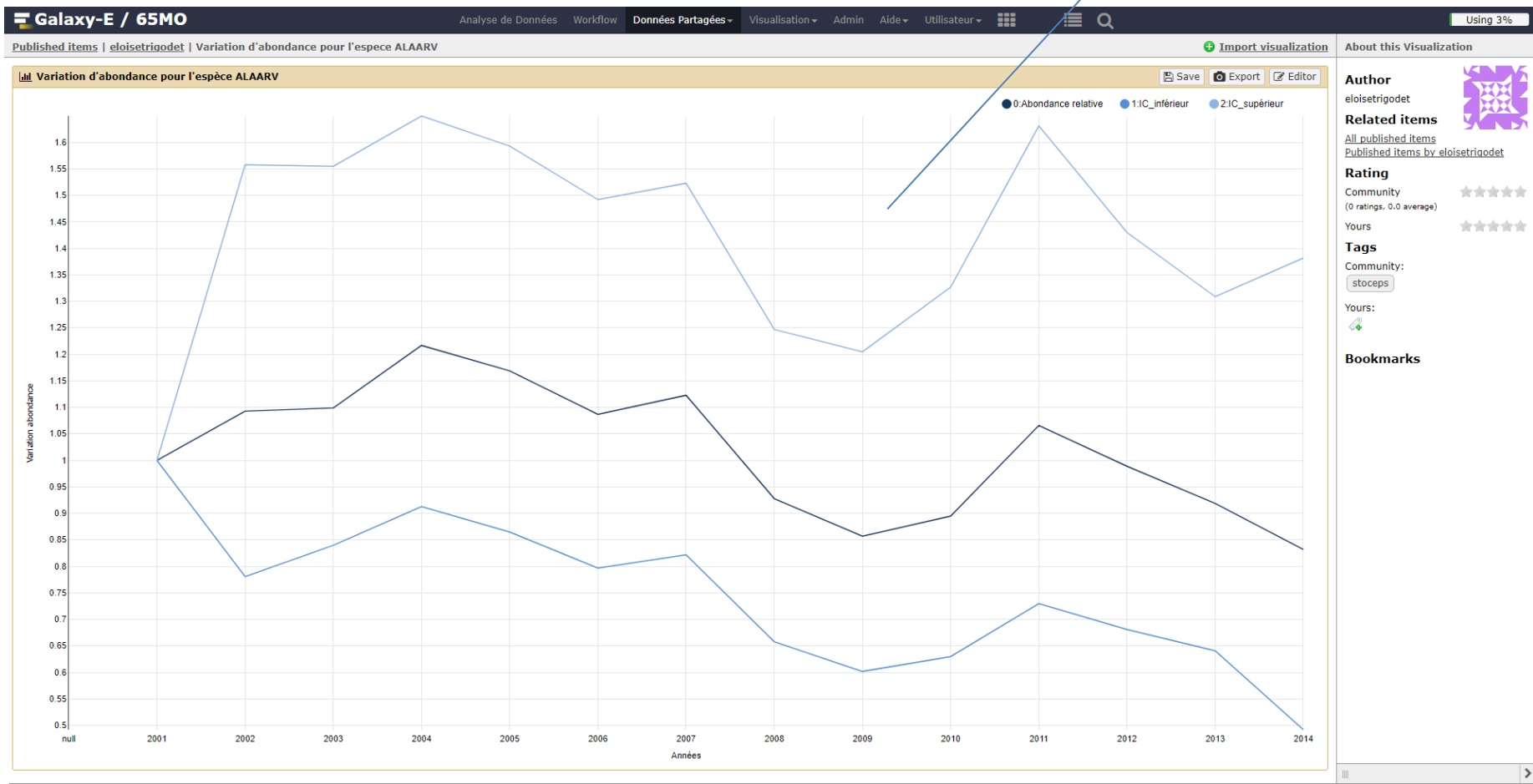
On the right side, there is a 'History' panel showing a list of data jobs. The top job is 'imported: Test Tadarida' with 4 shown and 12 deleted. Below it are two jobs related to '20160730_005625': '17: Trouver et Remplacer des patterns on da ta 3' and '3: output/20160730_005625'. The '3: output/20160730_005625' job is selected, showing 4 lines, 1 comment, and format 'csv, database: 2'. Below this, there is a 'randomForest 4.6-12' job with format 'Type rfNews() to see new features/changes/bug fixes'. At the bottom of the history panel, there is a table with columns 'Filename', 'CallNum', 'Version', and 'fileOur'. The table contains 4 rows of data for 'dataset_533.dat'.

At the bottom of the screenshot, there is a 'Visualisation de données tabulées' (Tabular data visualization) section. The URL at the bottom of the browser window is 'openstack-192-168-100-44.genouest.org/plugins/visualizations/charts/show?dataset_id=3da459a26c8f00b4'.

Visualisation de données tabulées

Représentation graphique simple

Exemple de représentation pour STOCeps



Parcours des données via graphes

Panneau d'outils

Ensemble des étapes de filtre / tri des données

Galaxy-E / 65MC

Analyse de Données Workflow Données Partagées Visualisation Admin Aide Utilisateur

Outils

recherche d'outils

Upload de fichiers
Manipulation de fichiers texte
Joindre, Soustraire et Grouper
Statistiques
Filtrer et Trier
Visualisation
Manipulation de collections
Tadarida
STOCeps
Workflows
• Tous les workflows

AskOmics

Select an entity to start a session:

Filter entities

Shared entities are *yellow* and personal entities are *blue*

Start

Or upload an AskOmics file

Load query

History

Rechercher des jeux de données

wf complet APECS sans regex
21 shown, 40 deleted
5.5 MB

10: Count on data 5
Mois

9: Count on data 4
Année

8: Count on data 4
Mois

7: Sort on data 4
Observation test

6: Filter on data 3
Echouage

5: Filter on data 3
Capture

4: Filter on data 3
Observation

3: Filter on data 2
>1997
2,523 lines
format: tabular, database: ?

Filtering with c5>1997,
kept 85.55% of 2949 valid lines
(2949 total lines).

Charts
Scatterplot
AskOmics
Ethercalc
Jupyter
openrefine
RStudio

2: Join two on data 1

1: Join two Datasets on data 3 and data 2

Invocation d'Askomics sur le jeu de données « 3: Filter on data 2 »

Parcours des données via graphes

Panneau d'outils

Ensemble des étapes de filtre / tri des données

The screenshot displays the Galaxy-E / 65MC interface. On the left, a sidebar contains a 'Panneau d'outils' (Tools Panel) with categories like 'Upload de fichiers', 'Manipulation de fichiers texte', 'Statistiques', 'Filtrer et Trier', 'Visualisation', 'Tadarida', 'STOCeps', and 'Workflows'. The main area shows a workflow graph with nodes for 'OrthologyGroup', 'Experiment', and 'Transcript'. A 'Launch query' button is at the bottom of the graph. To the right, a query editor for 'Transcript3' shows fields for 'end', 'start', and 'strand'. The 'end' field is set to '=', 'start' to '=', and 'strand' to 'minus'. On the far right, a 'History' panel lists workflow steps: 'wf complet APECS sans regex', '10: Count on data 5', '9: Count on data 4', '8: Count on data 4', '7: Sort on data 4', '6: Filter on data 3', '5: Filter on data 3', '4: Filter on data 3', '3: Filter on data 2', and '1: Join two Datasets on data 3 and data 2'. A dropdown menu is open over the 'AskOmics' step, listing visualization options: 'Charts', 'Scatterplot', 'AskOmics', 'Ethercalc', 'Jupyter', 'openrefine', and 'RStudio'.

Visualisation et interrogation du jeu de données par un graphe