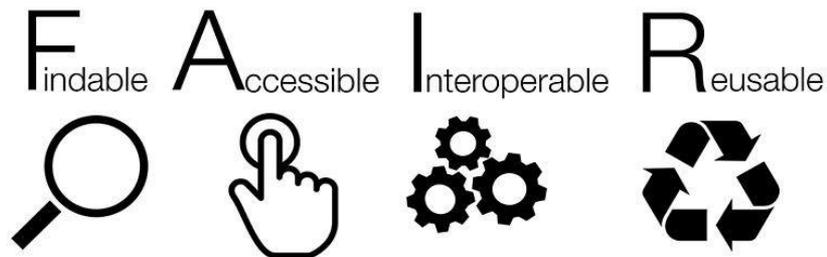




« L'écosystème » des données dans le domaine des sciences de la vie

Anne-Françoise Adam-Blondon

Séminaire de restitution de l'étude sur les données de séquençage de ressources génétiques pour l'agriculture et l'alimentation
8 octobre 2018



Contexte actuel des sciences de la vie

Open science ... open access, open data



"The days of keeping our research results to ourselves are over. There is far more to gain from sharing data and letting others access and analyse that data.

For example, if sharing big data reveals that a certain kind of cancer activates a particular molecular pathway in most cases and it turns out that there is already a drug approved and available to block the activation of that molecular pathway, clinical trials can begin almost immediately. Saving time, money and lives.

Or if scientists want to monitor the effects of climate change on local ecosystems, they can use Open Science to engage citizen reporting, and rapidly multiply the data at their disposal.

To make the most of Open Science opportunities for Europe, I plan to focus on open data, open access and research integrity over the course of my mandate."

*Commissioner Carlos Moedas,
"European research and innovation for global challenges", Lund, 4 December 2015*

<http://bookshop.europa.eu/en/open-innovation-open-science-open-to-the-world-pbKI0416263/>

Contexte legal autour de l'open data dans les services publics

<http://doranum.fr>

Les lois



Loi dite CADA

Relative à la liberté d'accès aux documents administratifs (dont indirectement aux données de la recherche) et à la réutilisation des informations publiques

1978

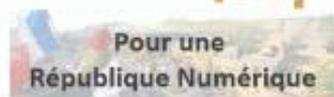
Loi dite Valter



Relative à la gratuité et aux modalités de réutilisation des informations du secteur public

2015

Loi dite Lemaire



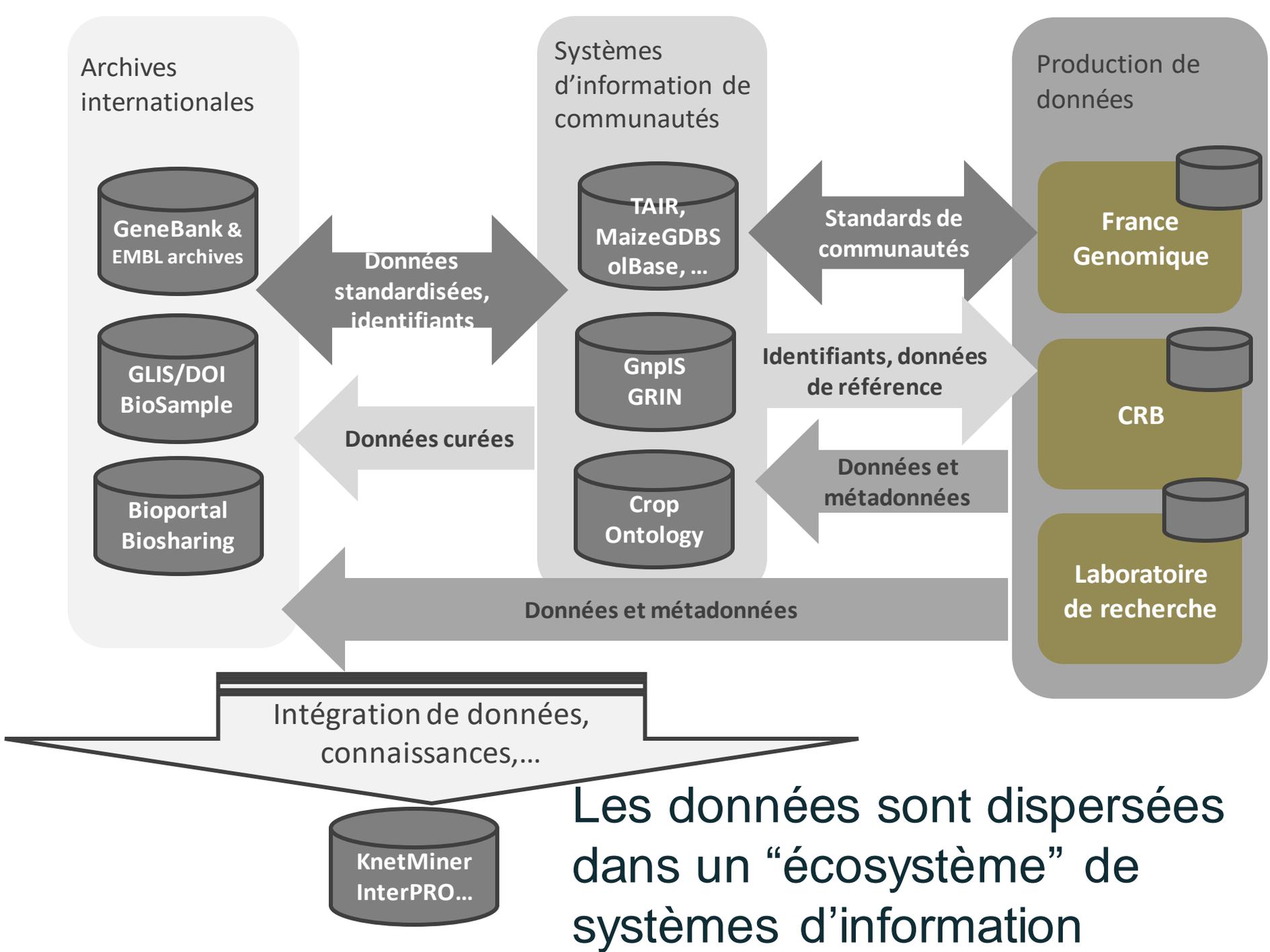
Relative à la circulation des données et du savoir, à la protection des droits dans la société numérique et à l'accès au numérique

2016

OPEN DATA Réutilisation
Conditions de Partage

L'Open Data trouve ses bases dans la loi dite CADA et c'est la loi dite Valter qui introduit le principe d'ouverture des données de la recherche.

La loi dite Lemaire crée un principe d'ouverture des données publiques par défaut. Les articles 30 et 38 posent un principe de libre utilisation des données de la recherche.



La centralisation des données n'est plus l'unique paradigme

- Déluge de données => les grandes archives internationales ne captureront volontairement probablement pas toute l'information (business model)
- Certaines données n'ont pas vocation à être en accès libre/archives internationales avec un accès complètement libre
- Pour certains types de données (eg. Phenotypes), la communauté internationale ne prévoit pas d'archive internationale à ce jour.

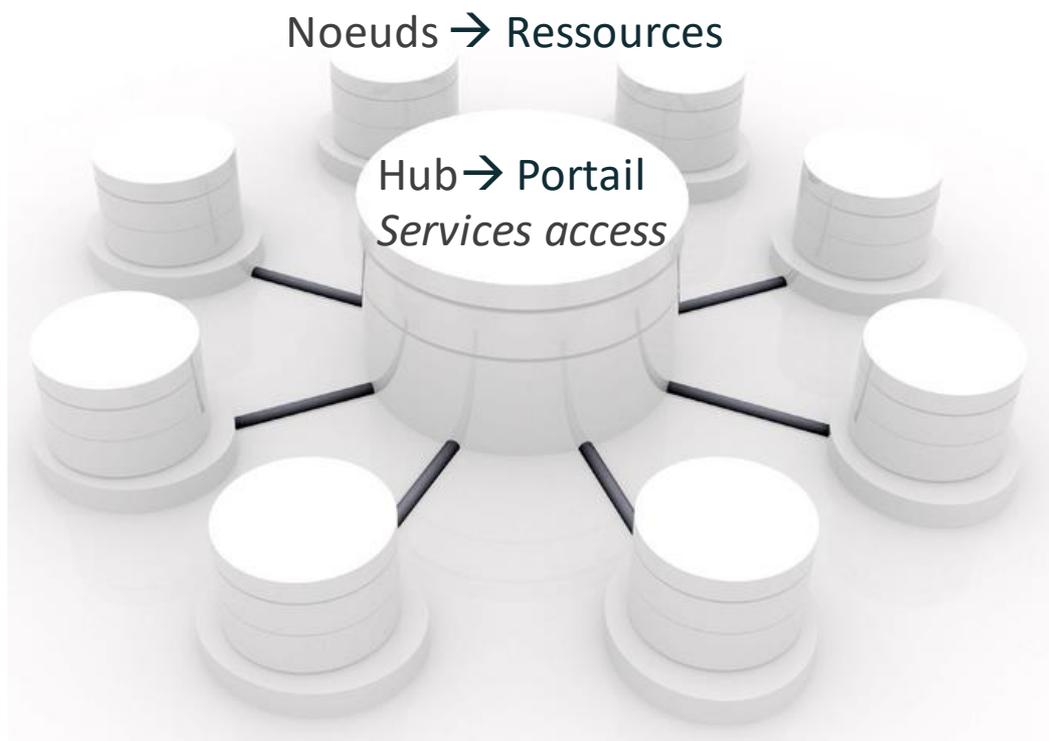
Emergence de fédérations de systèmes d'information

- Développement d'outils internationaux pour faciliter les flux de données au sein des e-écosystèmes: initiatives internationales (GA4GH, RDA, BreedingAPI, ...)
- Preuves de concept de ces fédérations: WheatIS, ELIXIR Beacons, fédération de la communauté plante d'ELIXIR, ...

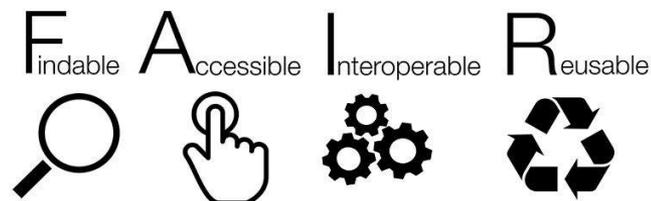


Organisation de ces fédérations

- Un réseau de nœuds (stables)
- Un portail central offrant des services (a minima: trouver et accéder aux données)
- Implémentation des principes FAIR



Recommandations développées par un collectif de bibliothécaires, établissement d'enseignement et de recherche pour l'Open Data en sciences



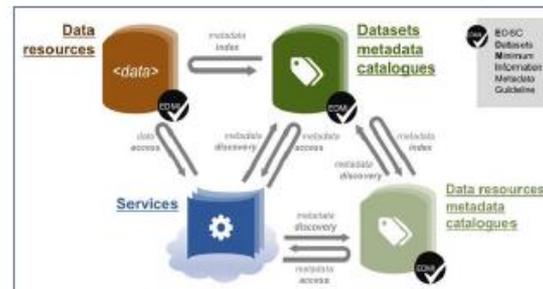
Wilkinson et al (2016) SCIENTIFIC DATA, 3:160018, DOI: 10.1038/sdata.2016.18

Organisation des e-Infrastructures européennes

- European Open Science Cloud: EOSC
- European Infrastructure of Bioinformatics for Life-sciences: ELIXIR

ELIXIR mission and strategy aligns with EOSC – implement life-science foundation in EOSC-Life

Deposition Database	Description	Infrastructure and identifier (Ecosystem)
Ensembl	Functional genomics data. Starts data from high-throughput functional genomics experiments.	European Bioinformatics Institute and the European Genome-phenome Archive
NCBI	Compendium of biological processes.	European Bioinformatics Institute and the European Genome-phenome Archive
ENA	Personalized human genetic and phenotypic data resulting from biomedical research projects.	European Bioinformatics Institute and the European Genome-phenome Archive
ENA	Multiple sequence information, covering raw sequencing data, collapsed and BWA-sequence assembly information and functional and taxonomic annotation.	International Nucleotide Sequence Database Collaboration
EMBL	Initial archive is freely available, open-access database underpins and provides basis for metadata links to Ensembl.	The International Nucleotide Sequence Database Collaboration
EMBL	Manufacture structures and their reference sequences as well as their biological roles, locations and characteristics, and experimental data from public experiments.	The International Nucleotide Sequence Database Collaboration
EMBL	Biological macromolecular structures.	The International Nucleotide Sequence Database Collaboration
EMBL	Mass spectrometry-based proteomics & protein expression information (2D-gel and the proteome).	The International Nucleotide Sequence Database Collaboration



FAIR data management in the life sciences

[10.7490/f1000research.1114985.1](https://doi.org/10.7490/f1000research.1114985.1)

Gateway for User access and mechanism for exposing life-science services (via *ELIXIR Registries*)

Compatible Cloud / Workflows / Reference Data Set Distribution Service



Preuve de concept d'une fédération internationale: wheatis.org

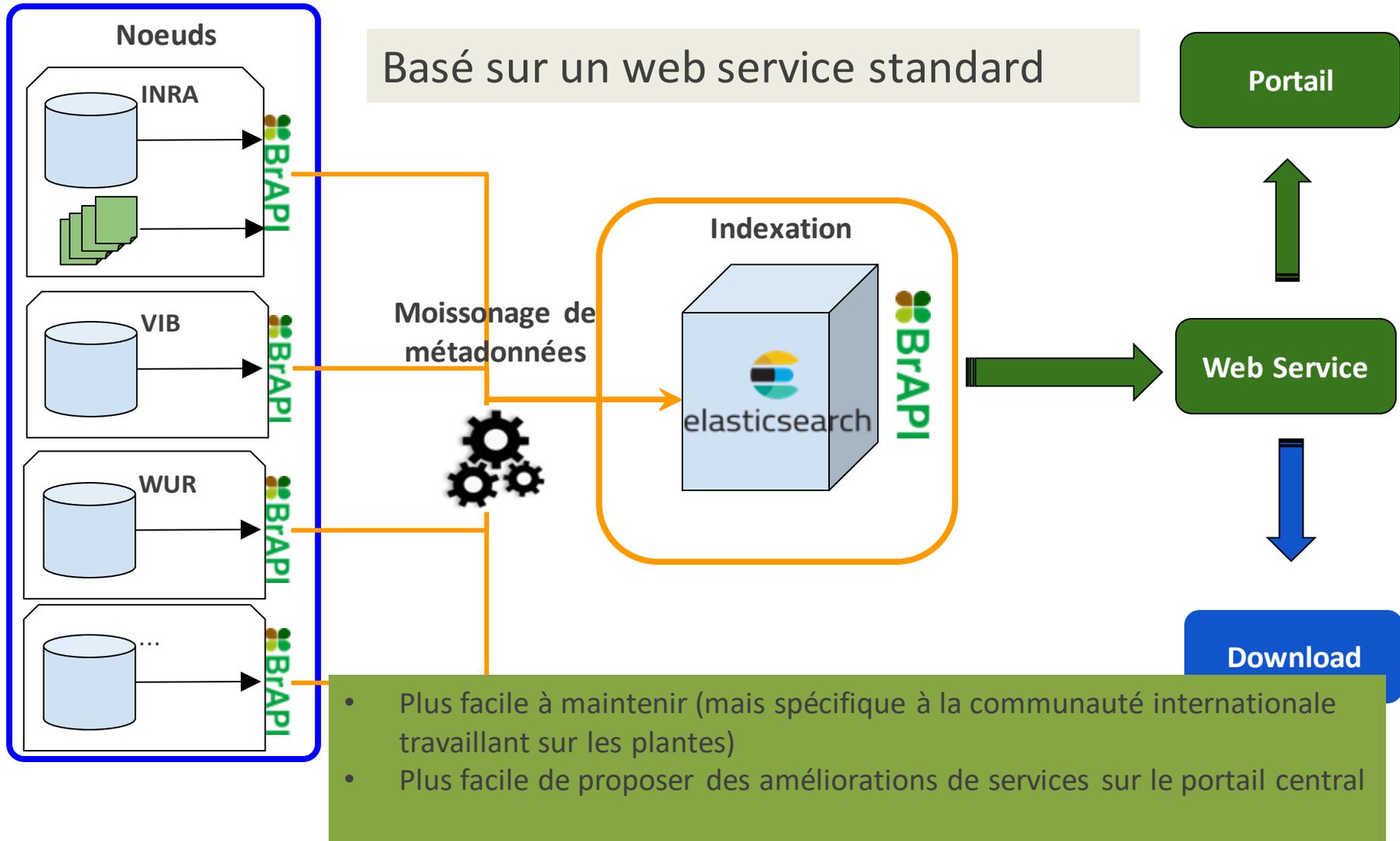
The screenshot shows the WheatIS website. At the top left is the 'WheatIS' logo. Below it are navigation links for 'Search' and 'About'. A section titled 'WheatIS nodes:' lists several international wheat research centers: transPlant-MIPS (UP) with CrowsNest: 13324; transPlant-IPK (UP) with CR-EST: 199220, GEBIS: 52878, and MetaCrop: 355; transPlant-EBI (UP) with Ensembl Plants: 218262; and transPlant-IGPAS (UP). To the right is the 'WHEAT INITIATIVE' logo. Below that is a search bar with the text 'Search in all WheatIS nodes' and a magnifying glass icon. Underneath the search bar are example search terms: 'Examples: fhb, wmc6530, Triticum, TRAES3BF001000010CFD'.



Fédération de SI de la communauté



ELIXIR



Remerciements

 H. Quesneville
C. Pommier
M. Alaux

 ist@inra
E. Dzale-Yeumo
S. Aubin

Financial supports



International infrastructures/initiatives



National and international crop projects



Pas que les plantes cultivées!

Merci!