

Retour d'expérience sur les étapes d'homogénéisation des données en écologie fonctionnelle végétale

Eric Garnier



5 juillet 2021

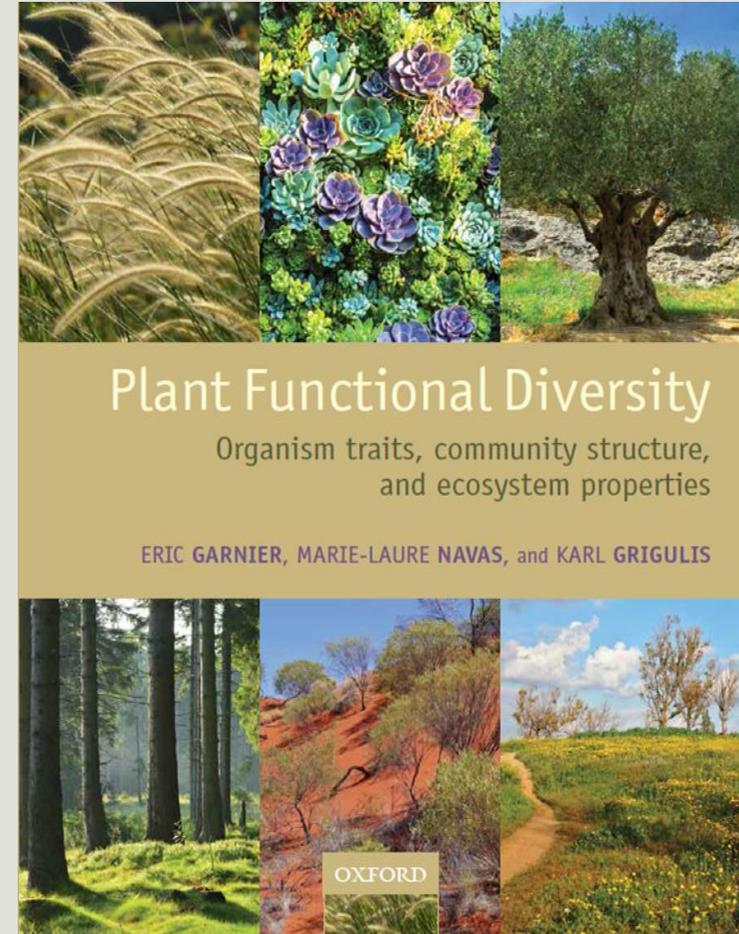
Plan de la présentation

- Introduction: écologie fonctionnelle et traits des organismes
- L'initiative TRY, base de données de traits des plantes : premières étapes d'homogénéisation des données
- Homogénéisation sémantique : le thesaurus des caractéristiques des plantes (TOP)
- Exemple d'application : identification de 2 dimensions fonctionnelles majeures chez les plantes

Ecologie fonctionnelle et traits des organismes

Approche qui présente un fort potentiel pour comprendre:

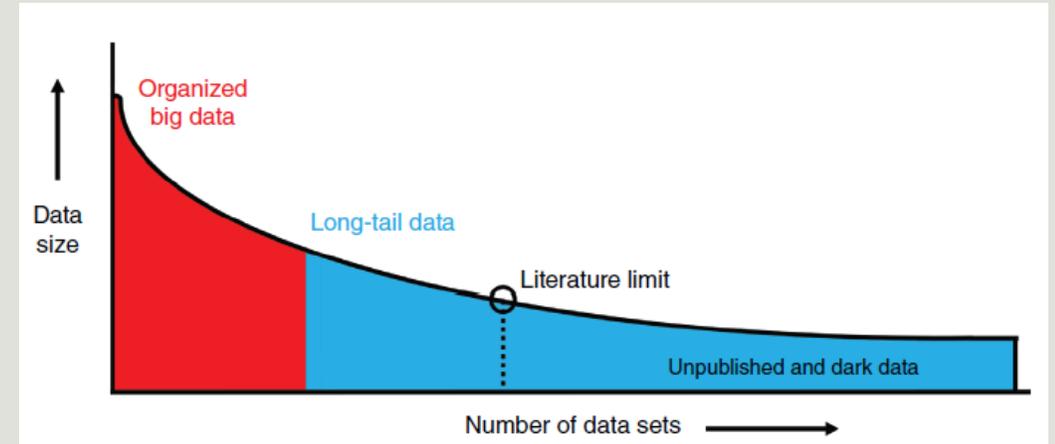
- les lois générales qui gouvernent le fonctionnement des organismes
- les relations entre biodiversité et facteurs du milieu (dont changement planétaires)
- les relations entre diversité et fonctionnement des écosystèmes
- les changements d'échelles entre niveaux d'organisation en écologie
- ...



Obstacles à l'interopérabilité des données dans le champ de l'écologie fonctionnelle

Techniques:

- Dispersion des données: petits jeux de données et sujets abordés très différents



Ferguson *et al.* (2014) *Nature Neuro* 17 : 1142

“Donnée obscure”: toute donnée qui n’est pas facilement découverte par un utilisateur potentiel
(Heidorn, 2008)

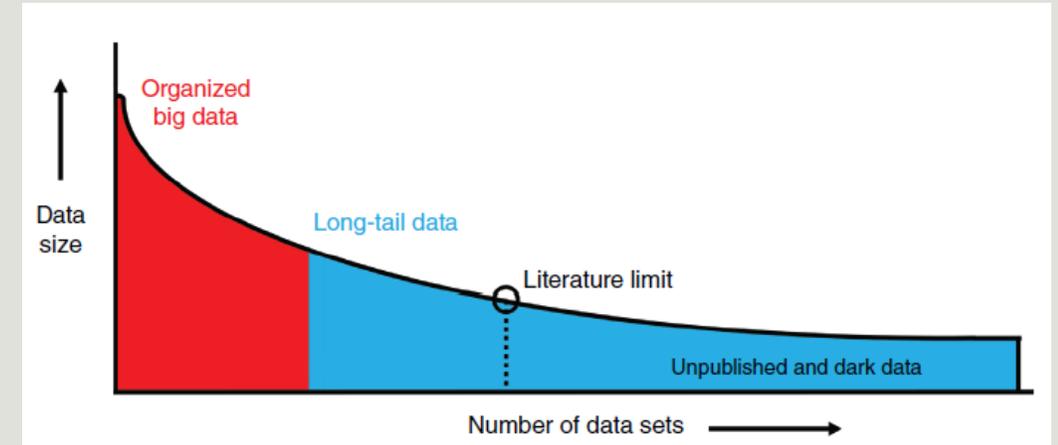
Obstacles à l'interopérabilité des données dans le champ de l'écologie fonctionnelle

Techniques:

- Dispersion des données: petits jeux de données et sujets abordés très différents
- Hétérogénéité syntaxique et sémantique des données
- Provenance des données: traçabilité, assurance qualité de collecte à la publication

Socio-culturelles:

- Peu d'incitation au partage
- Peu de protection contre les prédateurs
- Manque de reconnaissance pour la collecte des données
- Déficit chronique de financements et de temps dévolus à la mise en forme et au stockage des données



Ferguson et al. (2014) *Nature Neuro* 17 : 1142

“Donnée obscure”: toute donnée qui n'est pas facilement découverte par un utilisateur potentiel
(Heidorn, 2008)

L'initiative TRY – Une base de données mondiale de traits des plantes

Jens Kattge

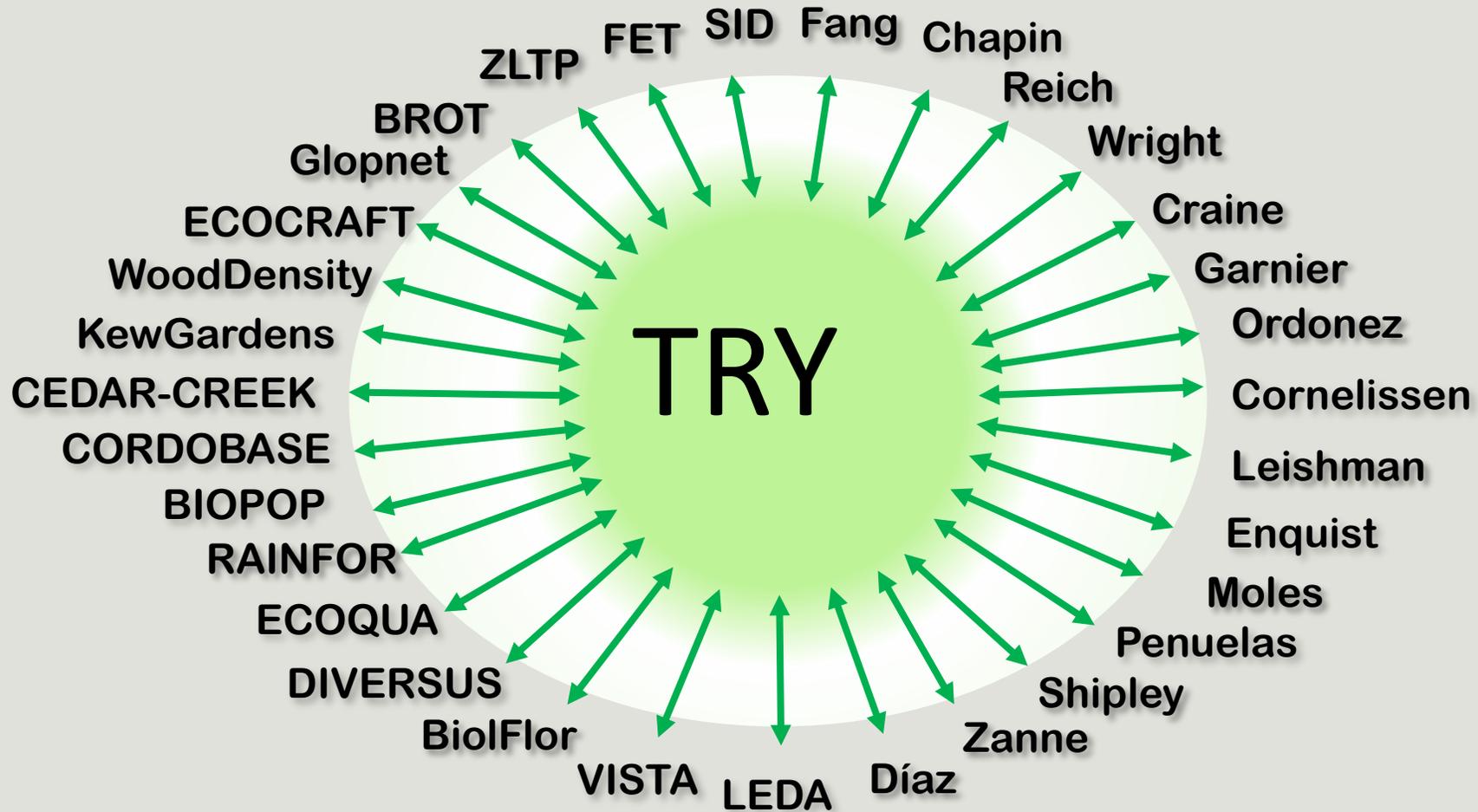
Christian Wirth Gerhard Bönisch Angela Nüske Sue Trumbore

Sandra Diaz, Sandra Lavorel, Colin Prentice, Paul Leadley
and all members of the TRY initiative



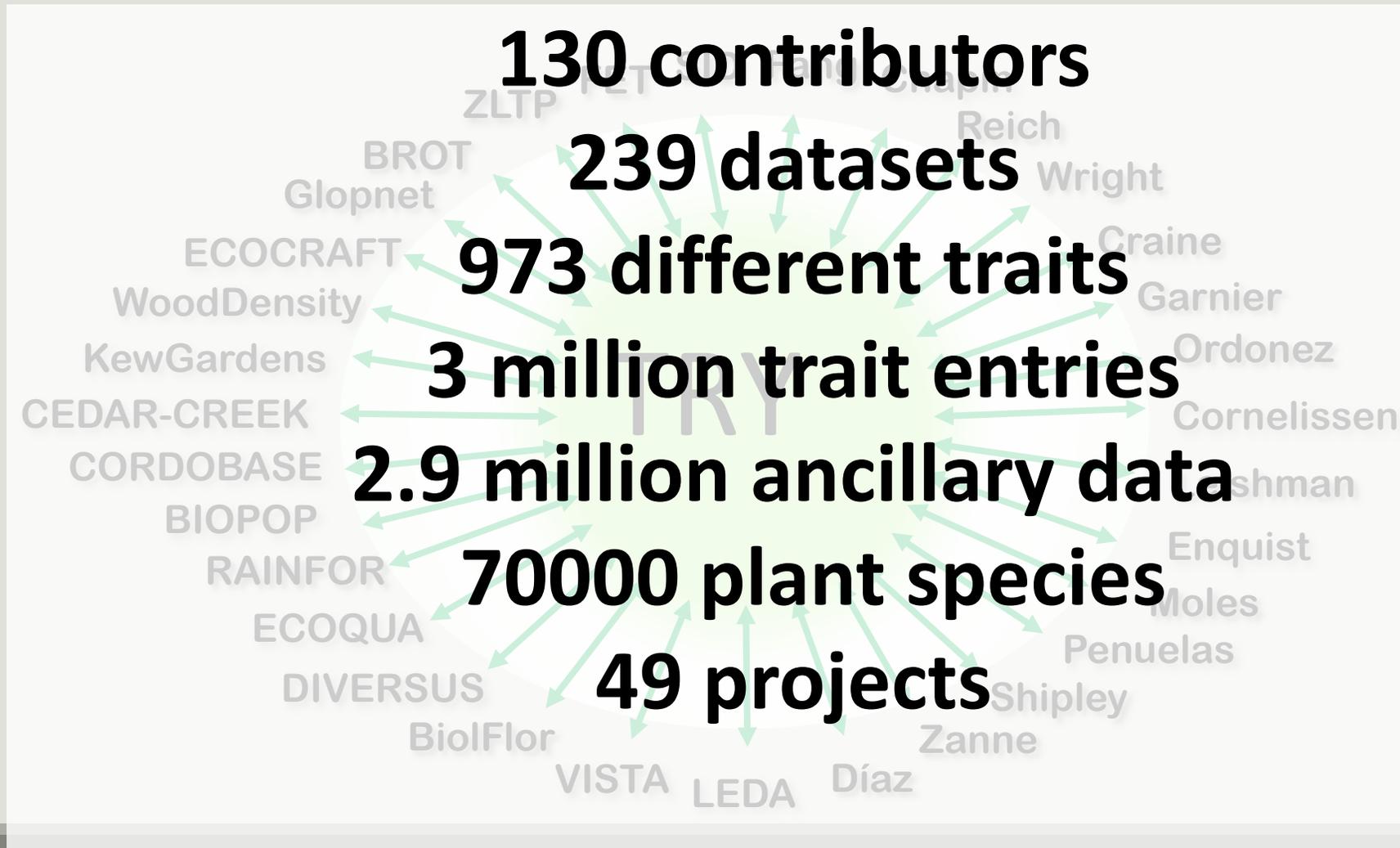
L'initiative TRY

seconde génération d'assemblage de données



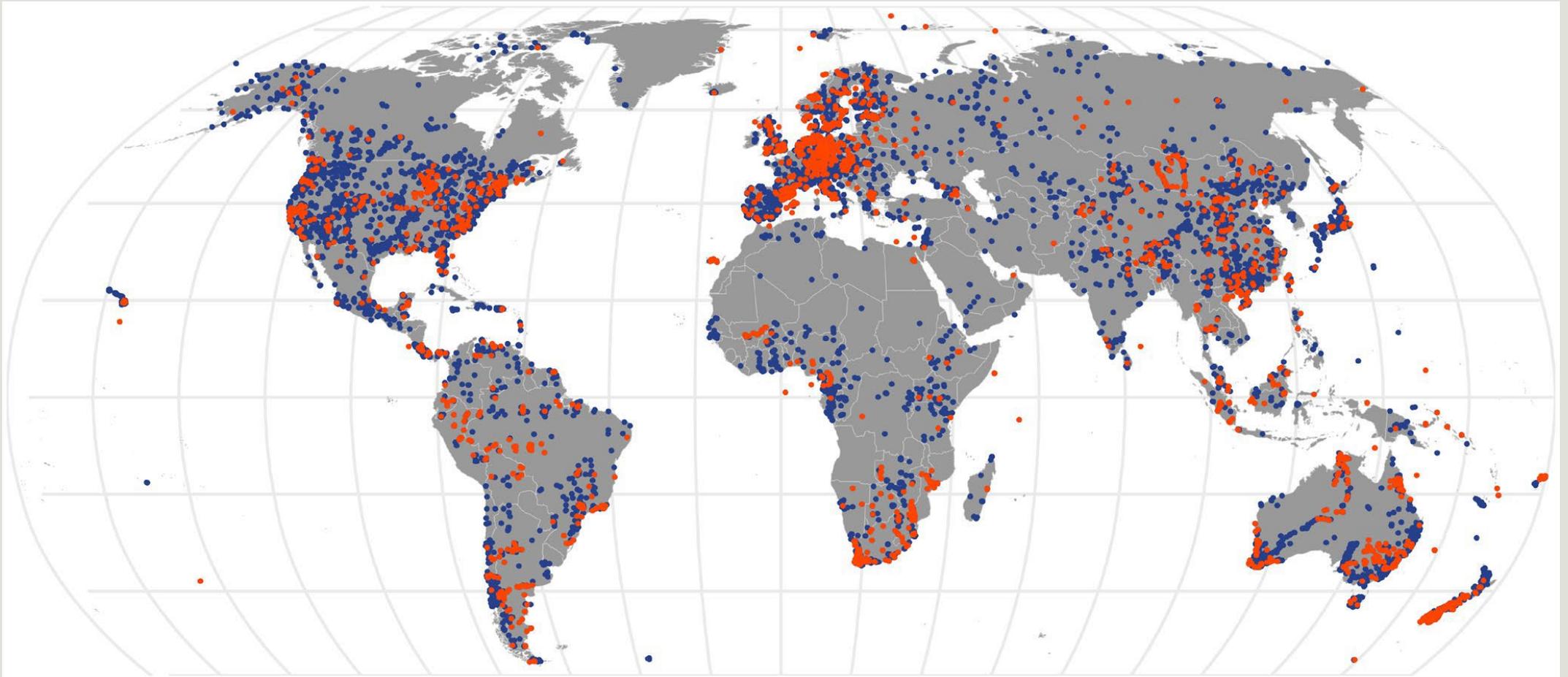
L'initiative TRY

seconde génération d'assemblage de données

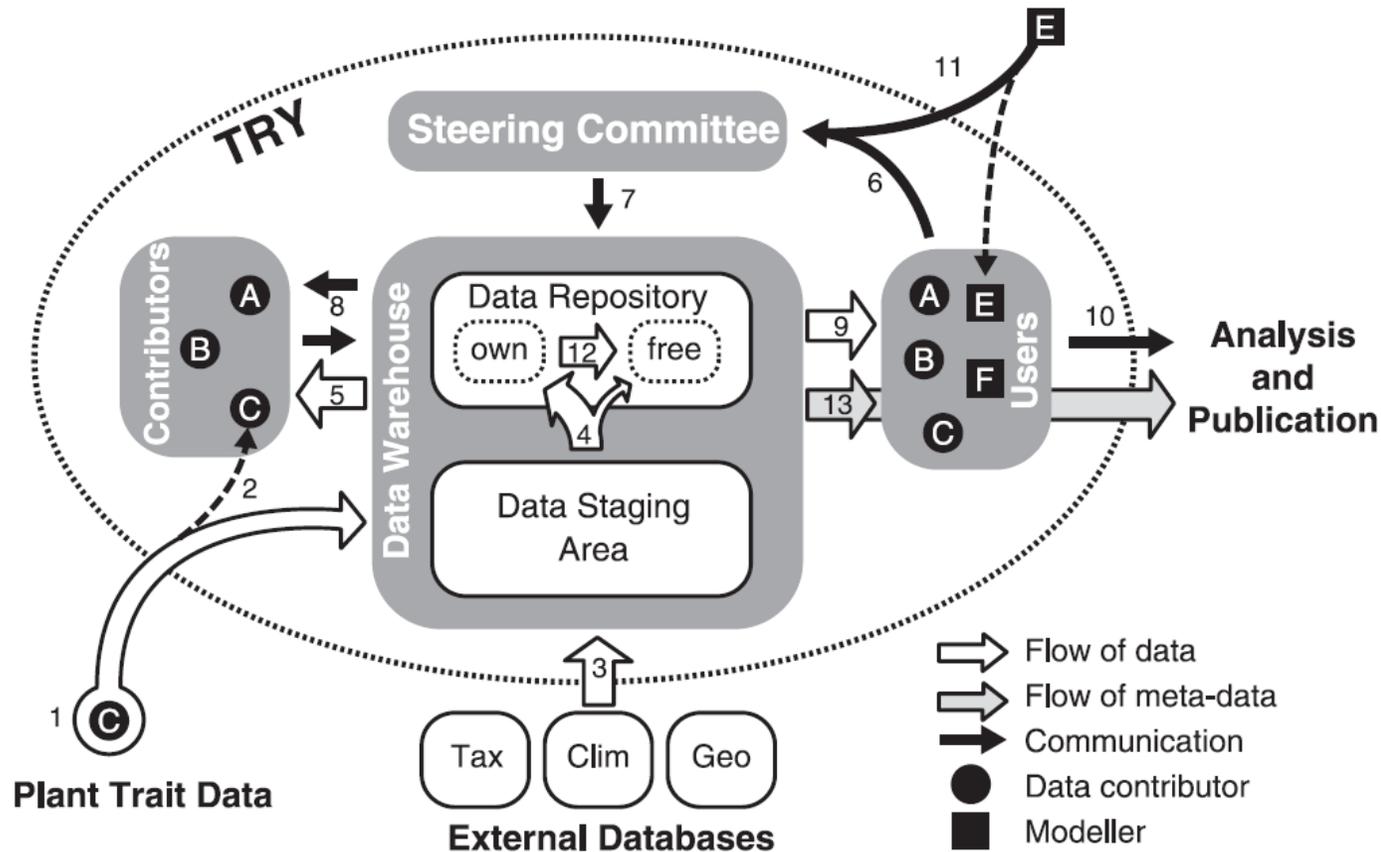


TRY, un réseau mondial:

> 200 participants, ~ 8000 sites de mesures



L'intégration des données dans TRY: premières homogénéisations



⇒ Homogénéisation syntaxique

⇒ Homogénéisation des unités

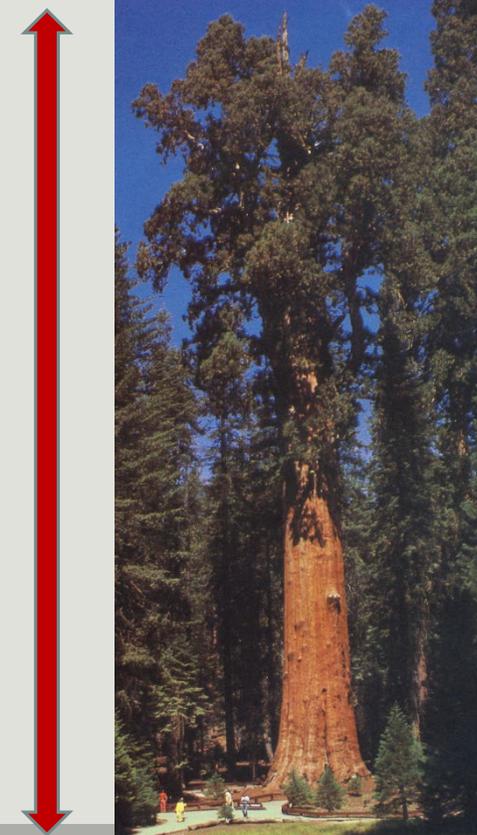
Nécessité d'une homogénéisation sémantique

Quelles sont les relations entre la **hauteur des plantes (plant height)** et les autres dimensions fonctionnelles?



LEDA: Kleyer *et al.*
(2008) *J Ecol*

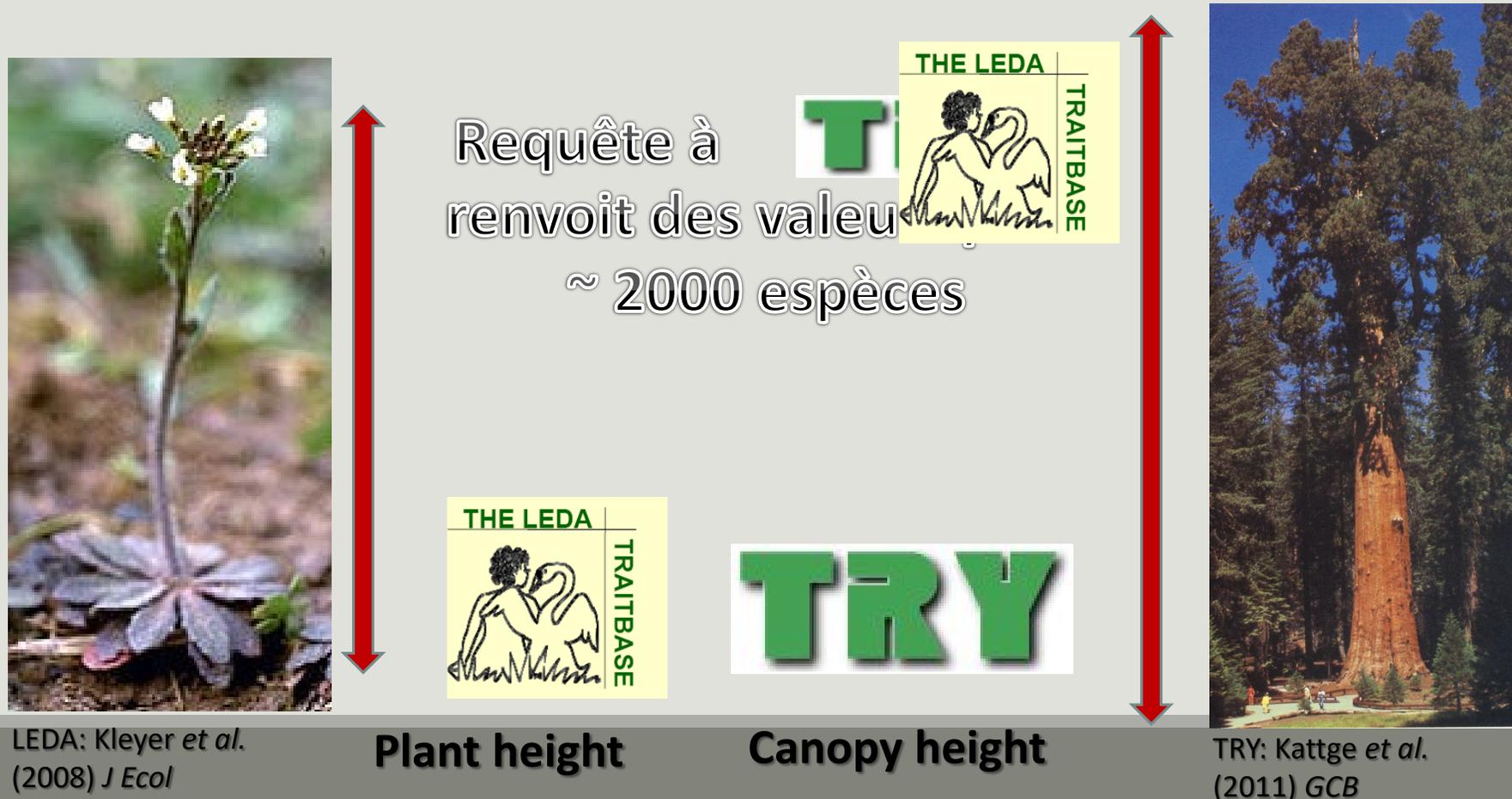
Requête à **TRY**
renvoit des valeurs pour
~ 2000 espèces



TRY: Kattge *et al.*
(2011) *GCB*

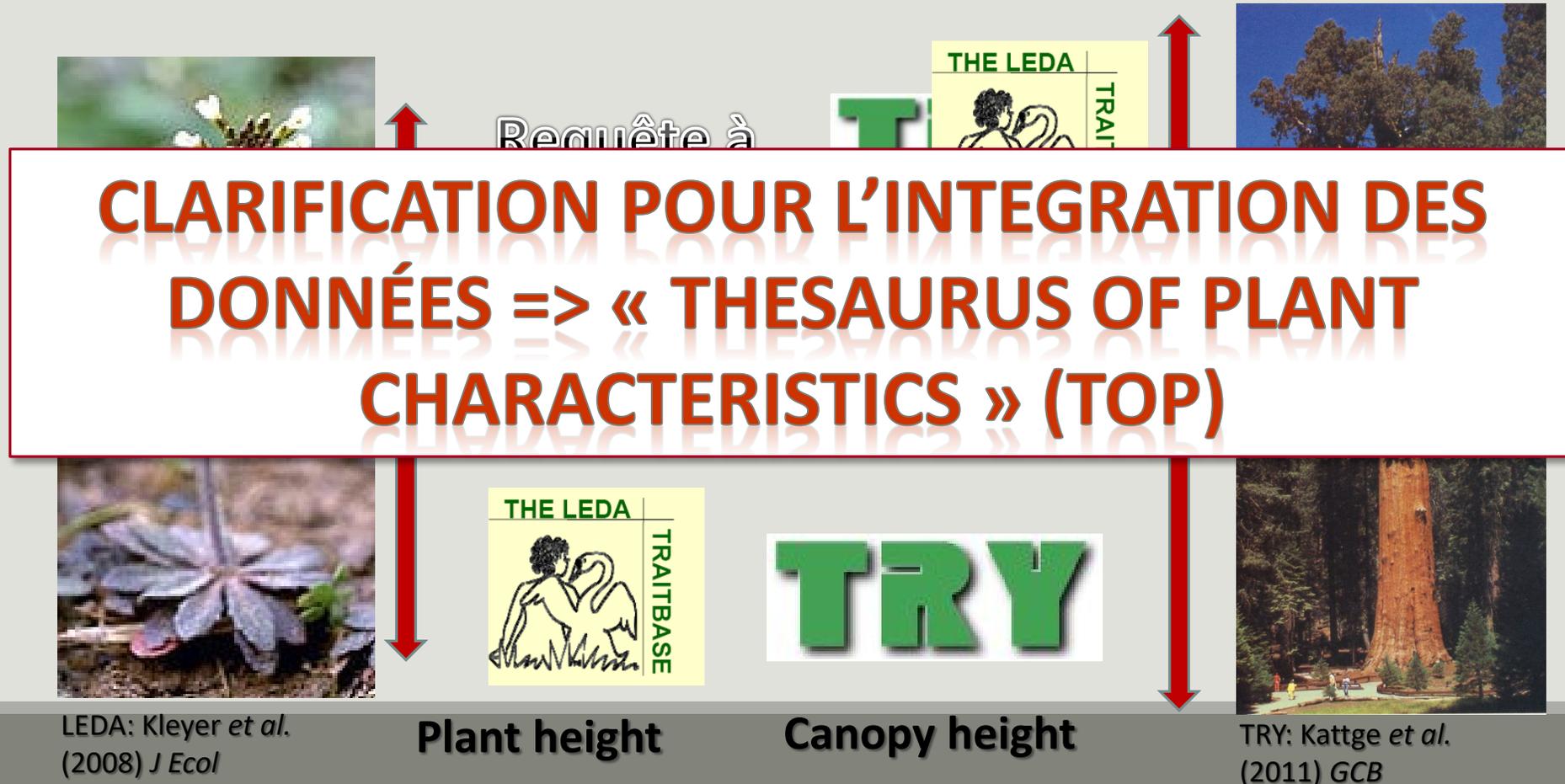
Nécessité d'une homogénéisation sémantique

Quelles sont les relations entre la **hauteur des plantes (plant height)** et les autres dimensions fonctionnelles?



Nécessité d'une homogénéisation sémantique

Quelles sont les relations entre la **hauteur des plantes (plant height)** et les autres dimensions fonctionnelles?



Fiche d'identité de TOP

- Ressource terminologique relative à une **approche fonctionnelle de la diversité végétale**, portant sur des concepts largement utilisés en écologie et évolution

- Présenté et décrit ici =>

Journal of Ecology 

Journal of Ecology 2017, **105**, 298–309 doi: 10.1111/1365-2745.12698

Towards a thesaurus of plant characteristics: an ecological contribution

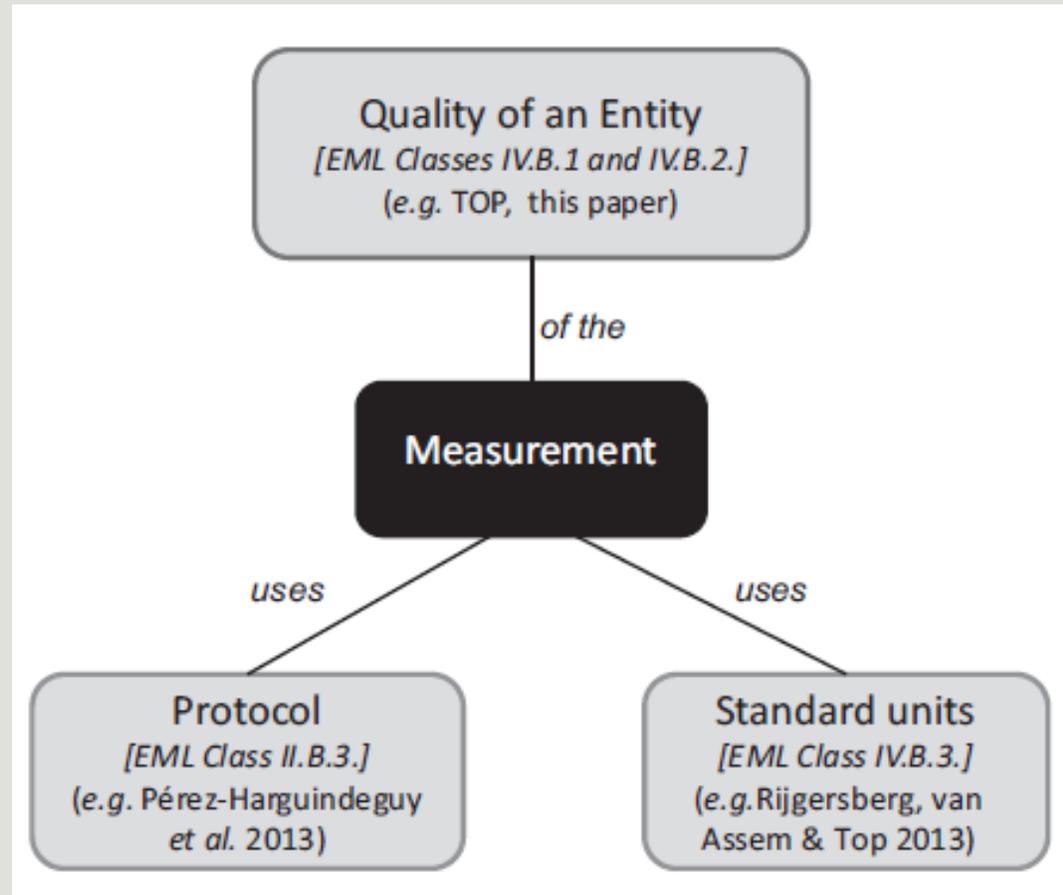
Eric Garnier^{*,1,2}, Ulrike Stahl^{3,4,5}, Marie-Angélique Laporte^{1,4,6}, Jens Kattge^{3,4}, Isabelle Mougénot⁷, Ingolf Kühn^{4,5,8}, Baptiste Laporte², Bernard Amiaud^{9,10}, Farshid S. Ahrestani^{11,12}, Gerhard Bönisch³, Daniel E. Bunker¹³, J. Hans C. Cornelissen¹⁴, Sandra Díaz¹⁵, Brian J. Enquist¹⁶, Sophie Gachet¹⁷, Pedro Jaureguiberry¹⁵, Michael Kleyer¹⁸, Sandra Lavorel¹⁹, Lutz Maicher^{20,21}, Natalia Pérez-Harguindeguy¹⁵, Hendrik Poorter²², Mark Schildhauer²³, Bill Shipley²⁴, Cyrille Violle¹, Evan Weiher²⁵, Christian Wirth^{4,26}, Ian J. Wright²⁷ and Stefan Klotz⁵

- Disponible ici =>
(et dans AgroPortal)

<https://www.top-thesaurus.org>



TOP est focalisé sur la sémantique: concepts et définitions



Identification de dimensions fonctionnelles des plantes à partir de l'extraction des données de TRY

Masse de la diaspore (SM)

Gamme: $3.0 \cdot 10^{-4} - 2.05 \cdot 10^7$ mg
 $n = 24,449$

Densité spécifique des tiges (SSD)

Gamme: $0.06 - 1.39$ mg mm⁻³
 $n = 11,356$



Hauteur de la plante (H)

Gamme: $0.001 - 90$ m
 $n = 24,720$

Surface d'une feuille (LA)

Gamme: $0.79 - 2.79 \cdot 10^6$ mm²
 $n = 12,173$

Masse surfacique feuilles (LMA)

Gamme: $4.9 - 1,507$ g m⁻²
 $n = 10,490$

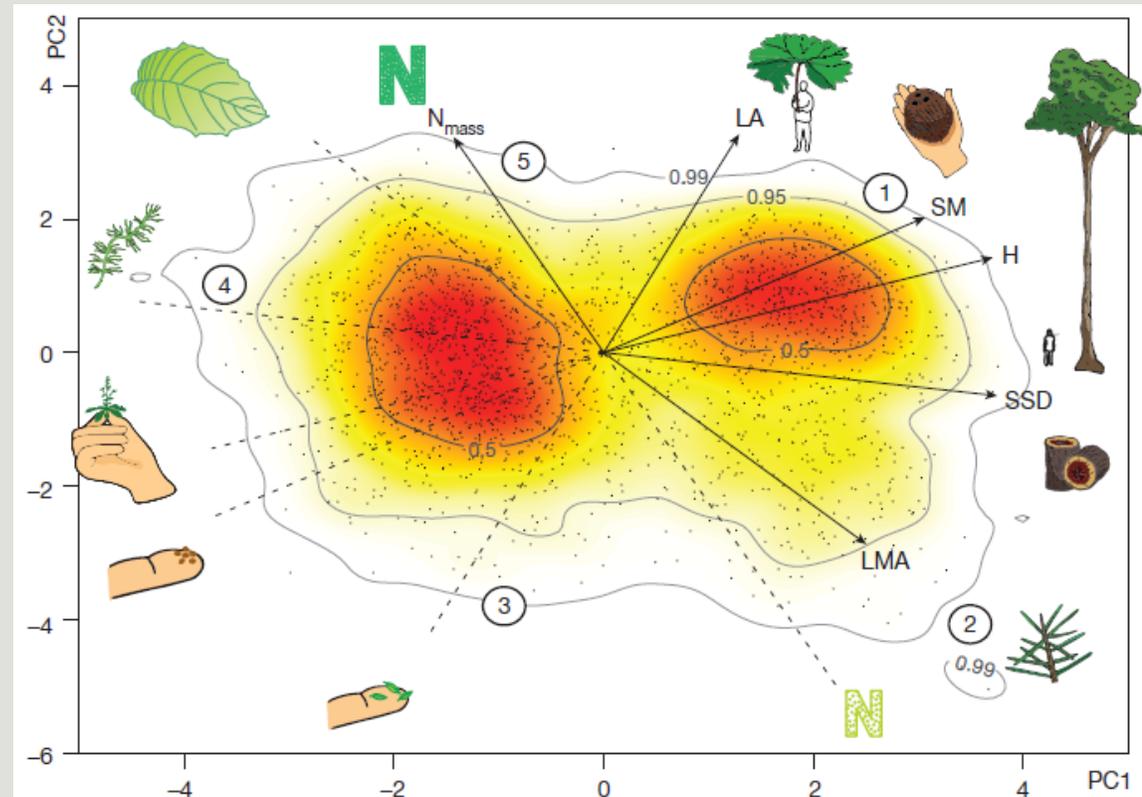
Teneur en azote feuilles (N_{mass})

Gamme: $2.48 - 68.98$ mg g⁻¹
 $n = 8,695$

Matrice complète pour 6 traits: 2,214 espèces

Le « global spectrum of plant form and function »

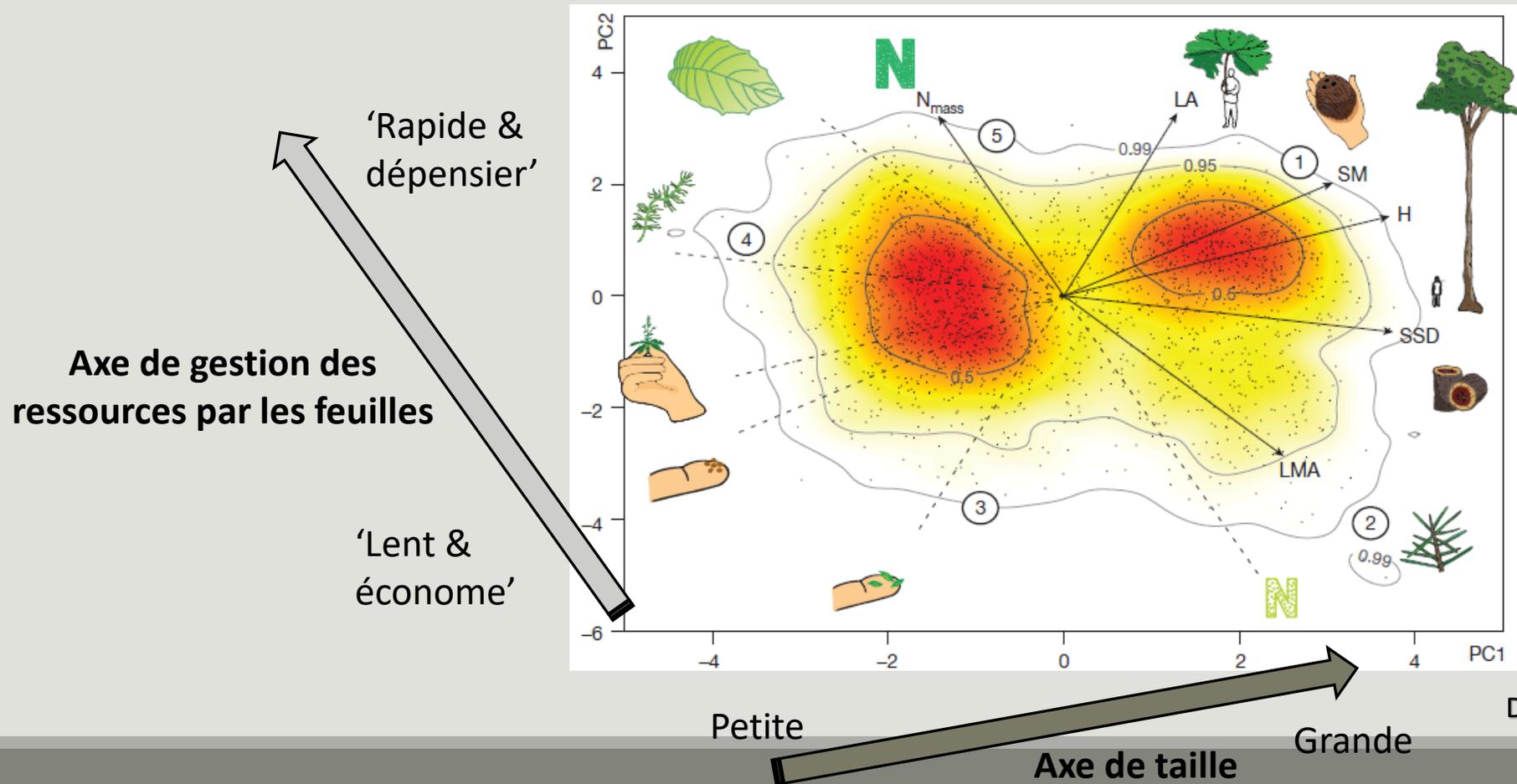
74% de la variation expliquée



Díaz et al. (2016) *Nature* 529: 167

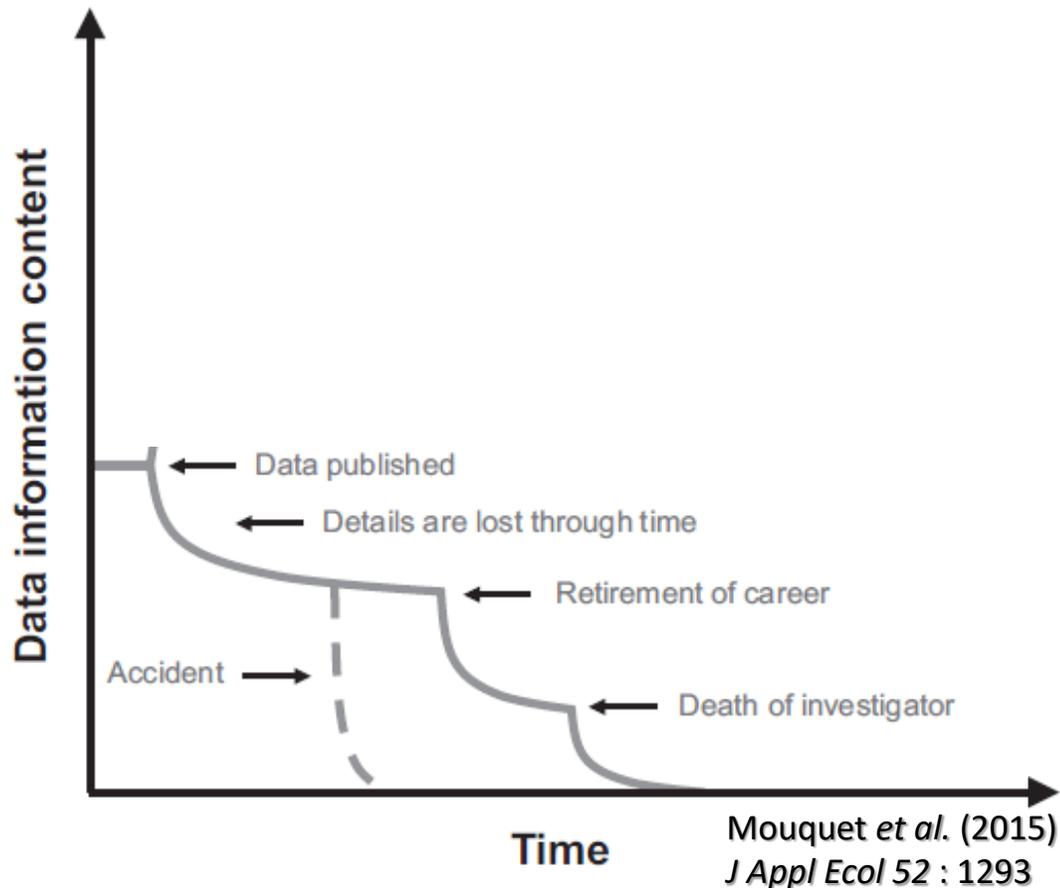
Le « global spectrum of plant form and function »

74% de la variation expliquée

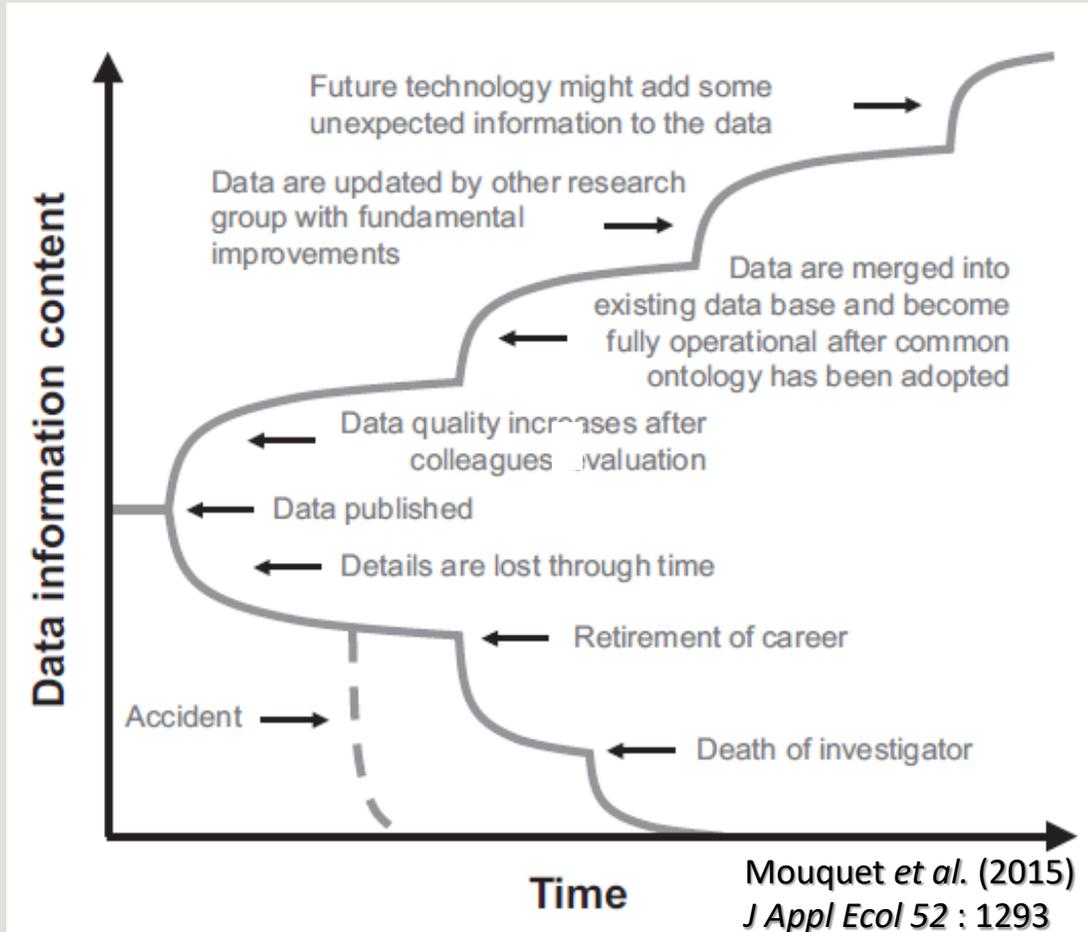


Díaz et al. (2016) *Nature* 529: 167

Pour conclure sur une note optimiste...



Pour conclure sur une note optimiste...



SCIENTIFIC DATA 

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.**

Findable **A**ccessible **I**nteroperable **R**eusable





Merci pour votre attention

Questions?