



CONTENT-MINING FOR TRANSPARENCY OF DRUG RESEARCH

@chris_kittel
@stefankasberger



SHUTTLEWORTH
FUNDED



1. open **bit.ly/cm-mozfest15**
2. open **pad**
3. copy files



1. Collaboration
2. Reproducibility
3. Big scale
4. Open Research Data
5. Do it together



Agenda:

1. Introduction (30min)
2. Hands-On (60min)
3. World Cafe (45min)



Introduction Round



ContentMine





THE SCALE OF THE TASK

- ~ 27,000 peer reviewed journals*
- > 5,000 publishers
- ~ 3,000 new papers per day
- “costing” 15 Billion USD to publish
- Representing 500 Billion USD of research

*Ulrich's database:

<http://ulrichsweb.serialssolutions.com/login>



The right to read
is the right to mine.

Facts in context

daily IUCN endangered species news

ous in *P. largha*, it was clearly heterogeneous in *C. ursinus*, *E. jubatus*, and ***Histriophoca fasciata***. Not only canines but also incisors were relatively highly variable in most cases although the canines

Processor: species

Set: daily20150828

Author(s): Mieczyslaw Wolsan, Satoshi Suzuki, Masakazu Asahara, Masaharu Motokawa,

CV and SDL indicate that the dentition of *H. fasciata* is the most variable in size, that of ***Eumetopias jubatus*** is the least variable, and those of *C. ursinus* and *P. largha* are at intermediate levels

Processor: species



© Andrew Reding,
http://creativecommons.org/licenses/by-nc-nd/2.0/deed.en_GB

Eumetopias jubatus

<http://dx.doi.org/10.2305/IUCN.UK.2012.RLTS.T8239A17463451.en>

NOT EVALUATED
NE

DATA DEFICIENT
DD

LEAST CONCERN
LC



Species

328

Mieczyslaw Wolsan, Satoshi Suzuki, Asahara, Masaharu Motokawa,

The IUCN Red List of Threatened Species™ 2015-3

Enter Red List search term(s) OTHER SEARCH OPTIONS Discover

[About](#) [Initiatives](#) [News](#) [Photos](#) [Partners](#) [Sponsors](#) [Resources](#)



Spotted seal

Animal

The spotted seal, also known as the larga or largha seal, is a member of the family Phocidae, and is considered a "true seal". It inhabits ice floes and waters of the north Pacific Ocean and adjacent seas. [Wikipedia](#)

Scientific name: *Phoca largha*

Higher classification: *Phoca*

Rank: Species

[en.wikipedia.org CC By-SA](http://en.wikipedia.org)

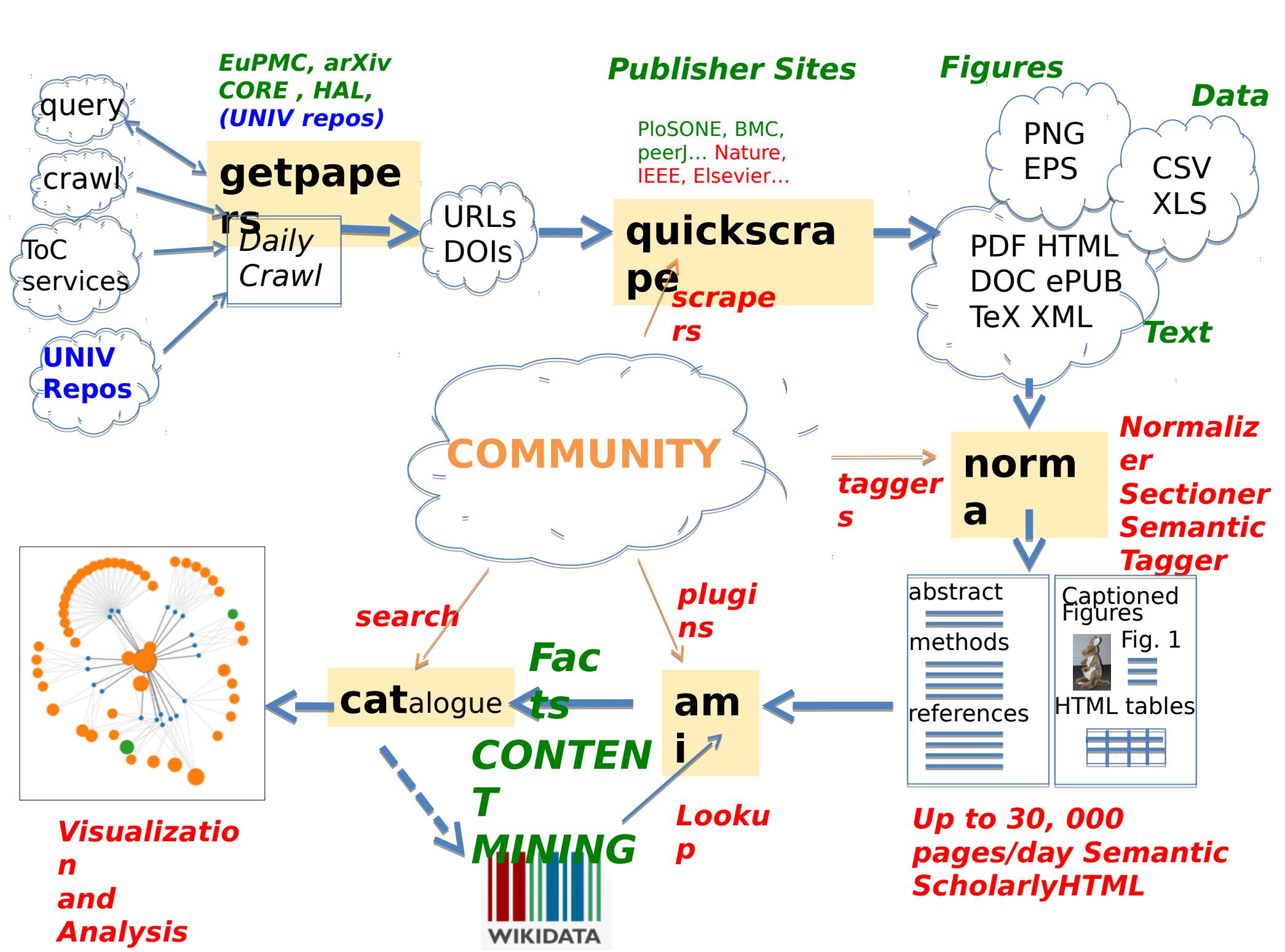


Phoca largha

<http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T17023A6730430.en>

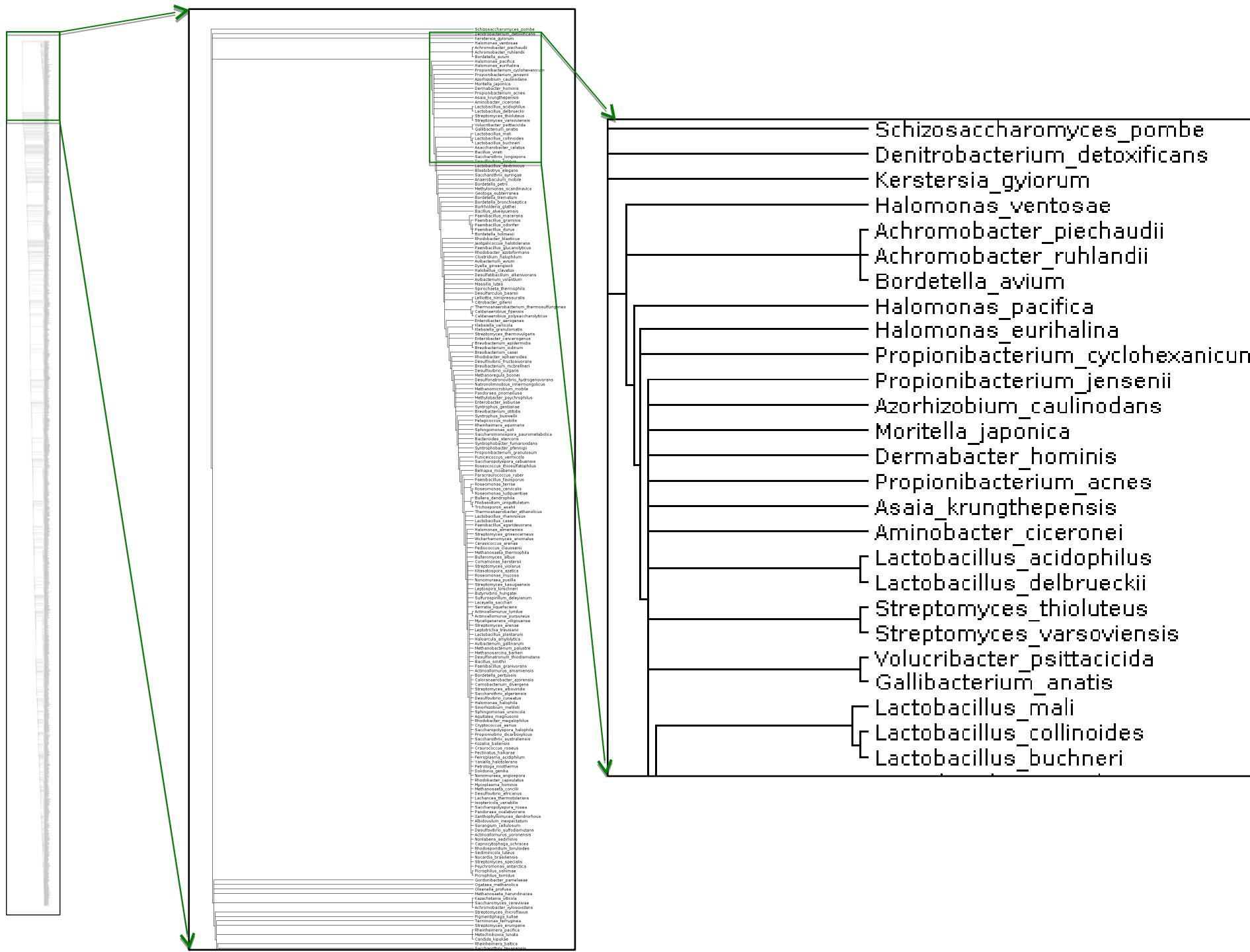


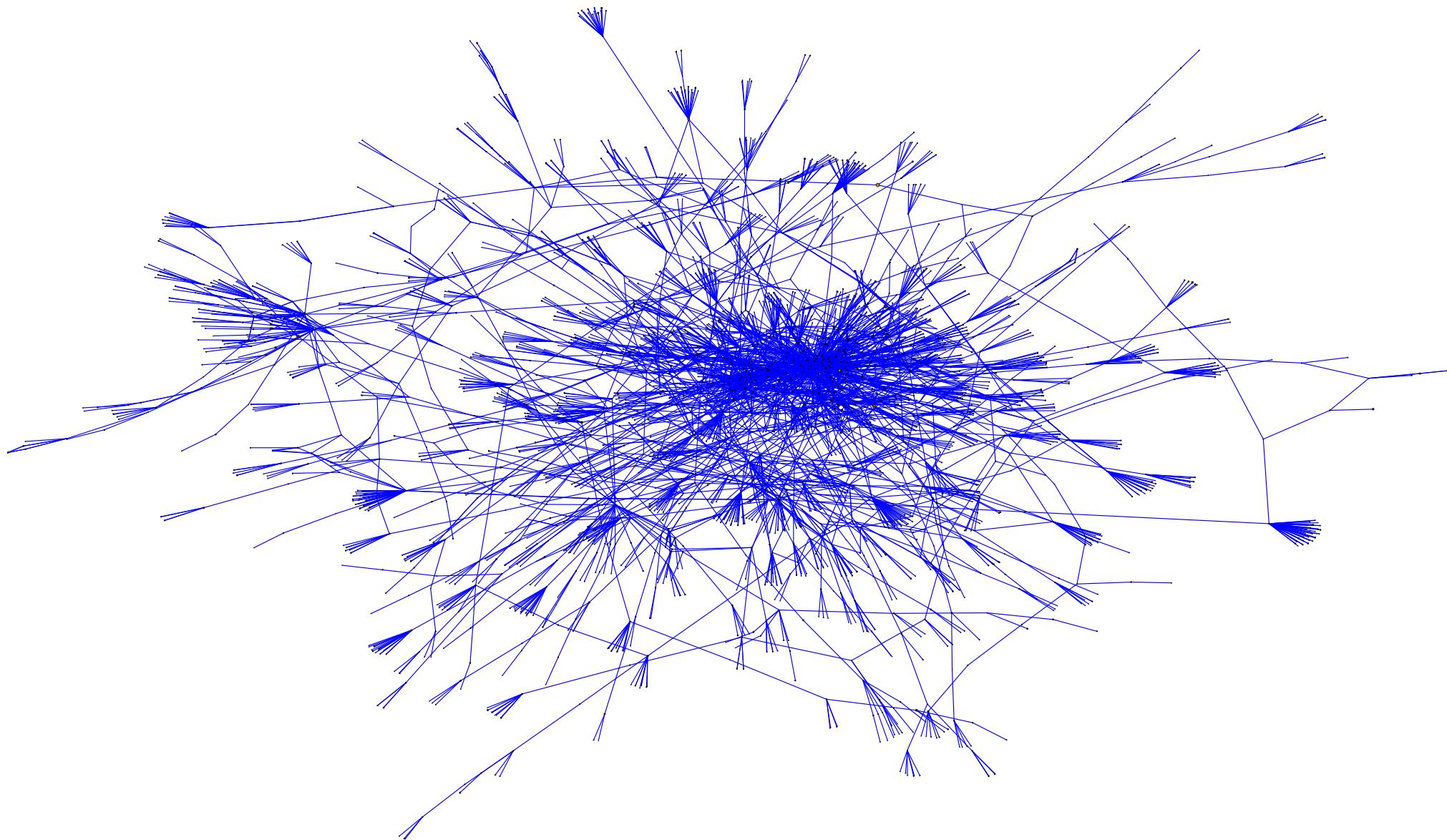
[Summary](#) [Classification Schemes](#) [Images & External Links](#) [Bibliography](#)



Supertree for 924 species

Tree







HACK WITH FACTS



What have you found?



WORLD CAFE



SHUTTLEWORTH
FUNDED



1. Get in groups (4-5 people)
2. 3 rounds (discuss and document)
3. Harvest in Circle



Questions:

1. What kind of questions could the data from the hacking session answer in terms of transparency and collaboration?

2. What are the opportunities you see in ContentMining on a massive scale? Think big!

3. What challenges do you see for ContentMining?



The right to read is the
right to mine.

contentmine.org



SHUTTLEWORTH
FUNDED