



PRE-PUBLICATION DATA SHARING: FROM SCIENTISTS' AND COMMUNITY PERSPECTIVE



Claudia Kasper

19 August 2020

SABRE-TP

Agroscope

1



What is pre-publication data sharing?

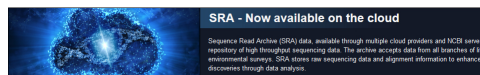
- archiving original data at publication is standard
- requested by scientific journals and funding agencies (e.g. SNSF)
- genomic data: raw sequence data; repositories e.g. ENA
- pre-publication data: raw data at the time of obtaining it, well before publication – «quick and dirty»

Open Research Data



Research data should be freely accessible to everyone – for scientists as well as for the general public.

The SNSF agrees with this principle. Since October 2017, researchers have to include a data management plan (DMP) in their funding application for most of the funding.



Agroscope

2

Why share data before publication?



- **rapid** (days-weeks-months vs years)
- **faster progress** of science
- avoid duplication of efforts
- use of data beyond project – **broaden utility**
- better use of public funds
- intensify collaborations between research groups
- **messy** (data curation not yet finished)
- **privacy/copyright** issues
- give up control of data
- **scooping**
- «data leeches»

Pre-publication data sharing can work...

...when players agree on rules - *Agreements*

Toronto Agreement (2009) – based on Bermuda (1996), Fort Lauderdale (2003) meetings

- **rapid prepublication data release** for large-scale genomics projects that have **broad utility** and are associated with **community buy-in**; creating reference data sets
- recommendations for funding agencies, **data producers**, **data analysts/users** and scientific journal editors

Toronto statement

Toronto International Data Release Workshop Authors, 2009
Nature Sep;461(7261) 168-170

- data producers
 - **state** their **intentions** and enable analyses of their data
 - **inform data users** about data being generated
 - provide relevant **metadata**
 - create **data bases**
- data users
 - freely analyse data and **act responsibly**
 - **respect**: allow data producers publish first global analysis of dataset
 - **cite** source accurately and completely
 - be aware of **data quality issues**
 - **contact data producers** to discuss publication plans

5

VarGoats Data Sharing Statement

http://www.goatgenome.org/vargoats_agreement.html

International Goat Genome Consortium

can serve as a model for pre-publication data sharing in Animal Genomics

Specifics:

- definition of terms beforehand (data, participant, analysis...)
- based on Fort Lauderdale and Toronto agreements
- members agree to statement
- private shared storage vs public data



6



Suggestions for pre-publication data sharing of pig genomic data CH

Goals: reduce time data collection – data availability to all Swiss pig genomics groups to accelerate research and make better use of existing data. Transparent and fair «consortium» instead of intransparent bilateral agreements.

- ~~rapid~~ → quick is good enough
- **participants:** ETH, HAFL, VetSuisse Bern, Suisag, Agroscope
- **participants have both roles** – data producer and data user
- **storage:** Swiss server, private access to participants, safe (backups). Ideally in public domain (Agroscope or KITZ)



Outlook

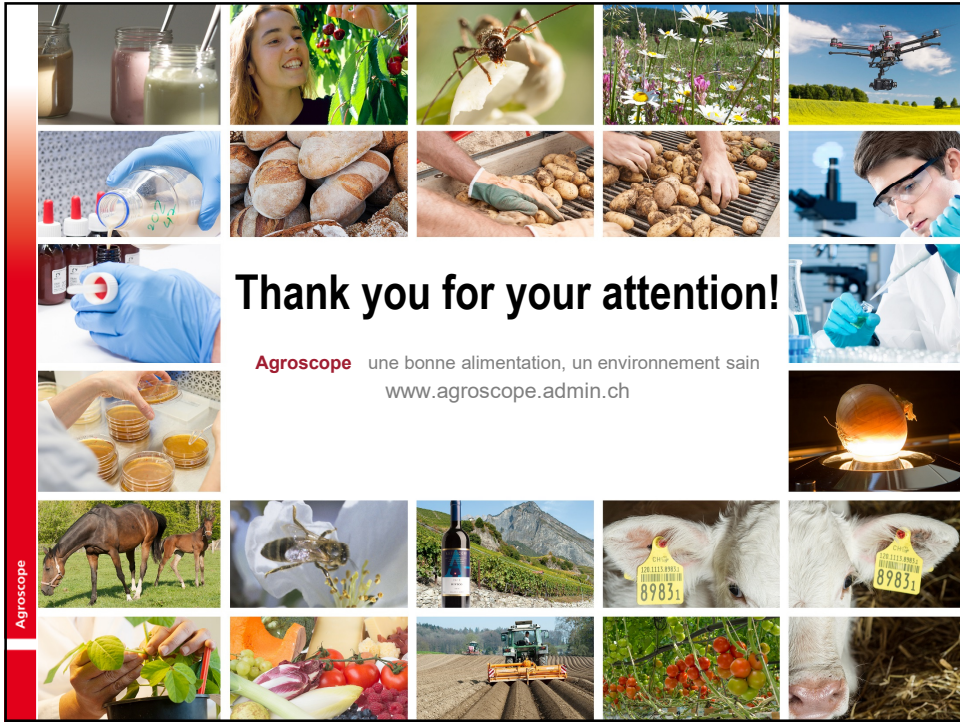


equal partnership between all players

clear



prevent misunderstandings, increase transparency



Thank you for your attention!

Agroscope une bonne alimentation, un environnement sain
www.agroscope.admin.ch

Agroscope