Disease data collection in aquaculture; how to collect and interpret data

Britt Bang Jensen,
Norwegian Veterinary Institute

21st Annual workshop of the National Reference Laboratories for Fish Diseases

Kongens Lyngby, May 30-31, 2017

Veterinærinstituttet
Norwegian Veterinary Institute

Why this talk?

- **Data = knowledge**
- **Data** is a set of values of qualitative or quantitative variables.
- Difficult to get good data
- Good data=Data fit for purpose

Fit for purpose:

- Formulating hypothesis is alpha-omega!
- Examples:
  - Surveillance
    - Population data and sample data required
  - Tracing outbreaks
    - Sequence data required
  - Modelling disease spread
    - «Negative» data required
  - Documenting disease freedom
    - Production data and geographic coordinates required
    - Informed idea of prevalence
  - Etc

Identify data sources
Disease data:

- National/regional screening programs
- National reference laboratories
- Private/industry initiated screening programs
- Private laboratories
- Medicine registers
- Insurance companies
- Projects with specific purpose for sampling

Norwegian Veterinary Institute

Identify data sources
Production data:

- National/official registers
- Aquaculture associations
- Slaughterhouses
- Questionnaires/interview
- Feed producers, boat handlers etc

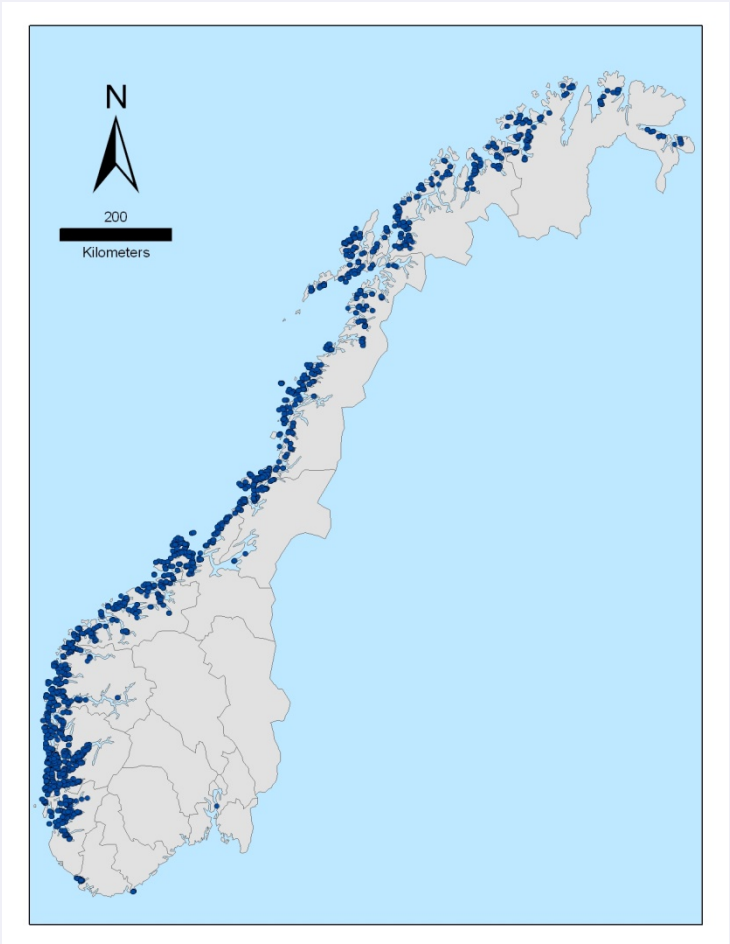
Norwegian Veterinary Institute

Data collection

- Logistic systems must be in place
  - to handle the data
  - for sharing obtained information with the competent authorities and the public.
- Different systems for data collection are available
  - preferable are those which are automated and require little work from the suppliers of data.

Official data:
Data from aquaculture register

- Geo-references
- Ownerships
- Production characteristics
  - Type
  - Species
  - Size
- (Category as required by EU-law)



Norwegian Veterinary Institute

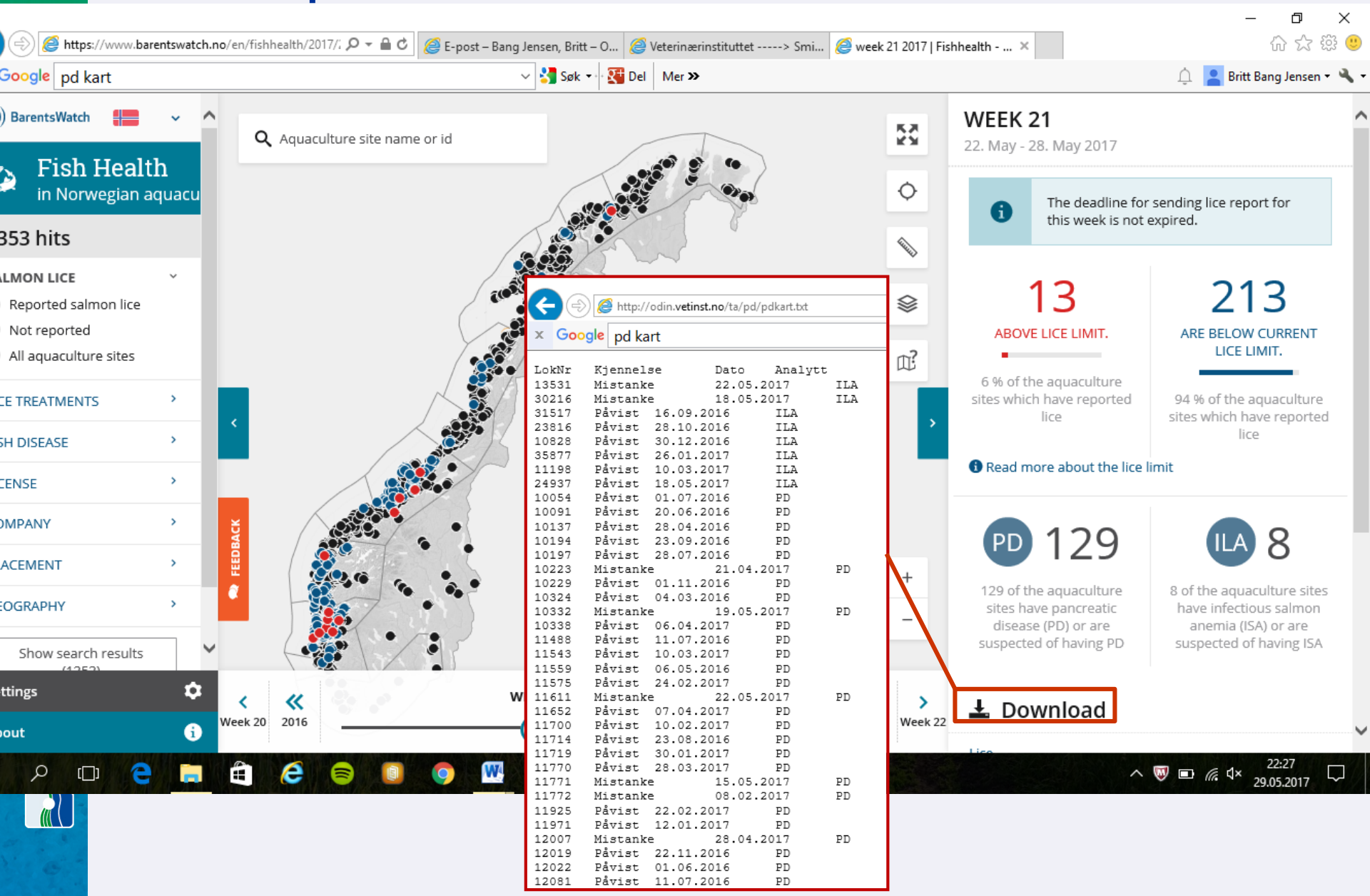
Official data: Laboratory/NRL data

- Notifiable diseases
  - Date for diagnosis
  - Species
  - Removal of stock
  - Subtype/genotype
- Non-notifiable diseases
  - If farmers agree
  - Screening data

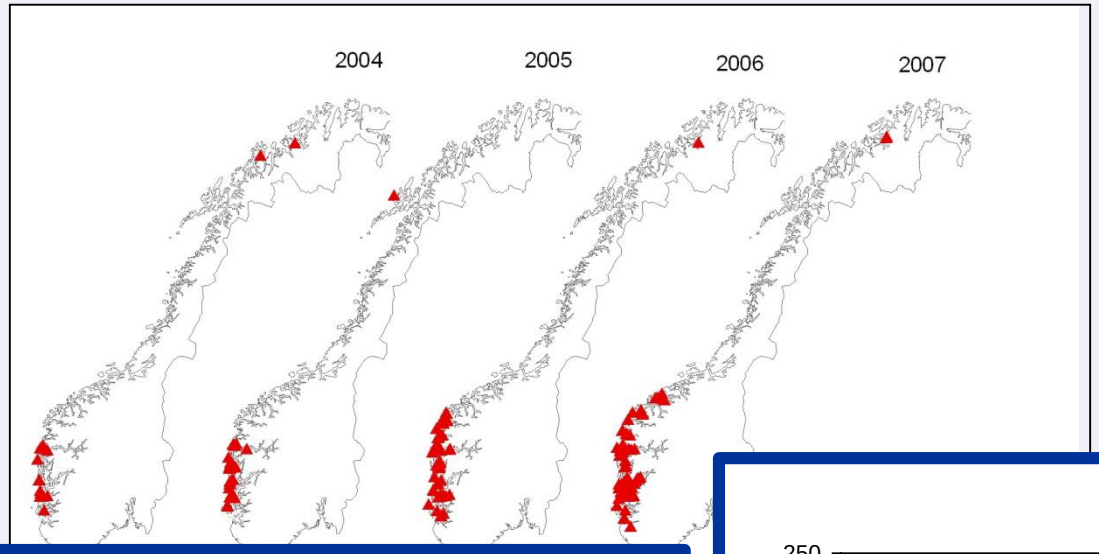
Considering disease data:

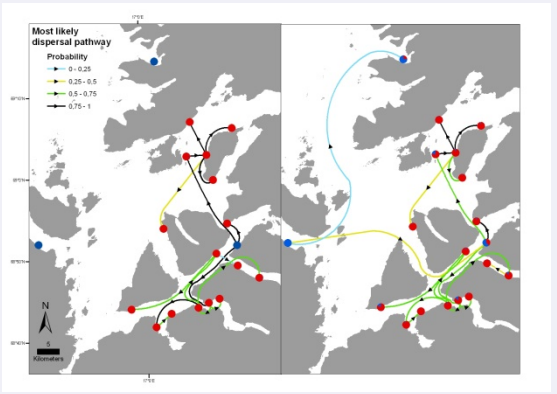
- Case definition:
  - PCR/histo/serology/patognomonic symptoms
  - Clinical disease?
  - «Accidental finding»/Disease investigation
- Positive cases
  - Time for diagnosis
  - Sequence
  - Geographical coordinates
- Negative locations
  - Screening/inspection program?
  - Known negative or unknown status?

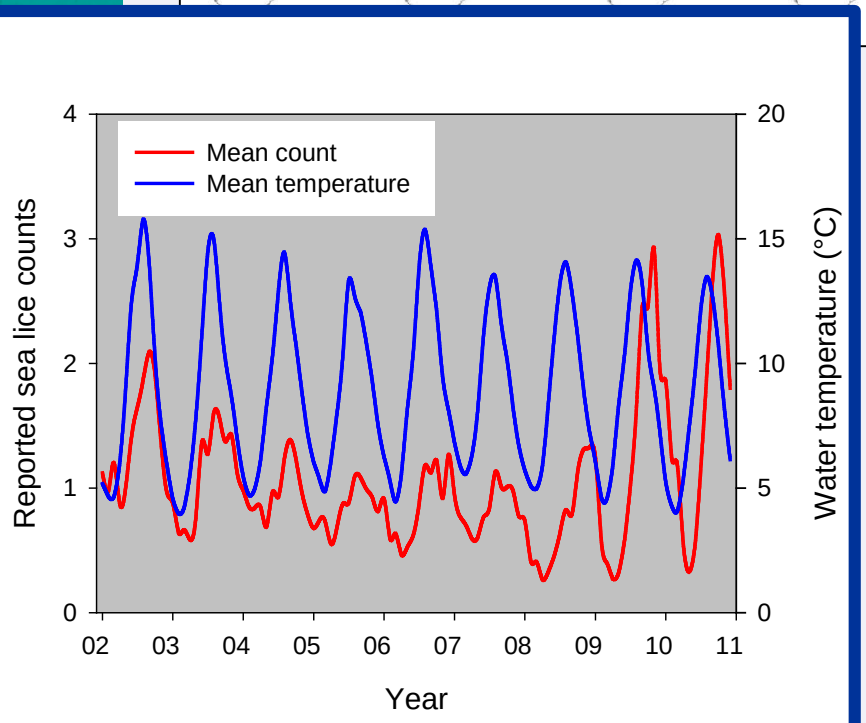
Examples of data use: www.barentswatch.no

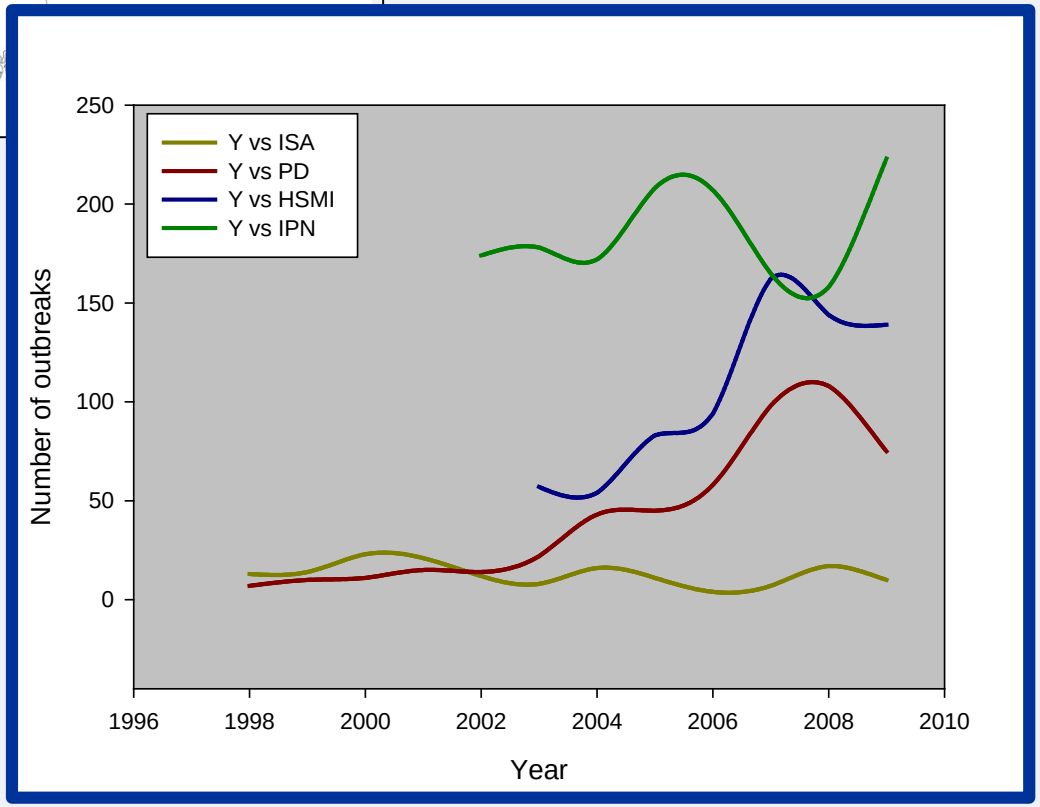


Disease development over time



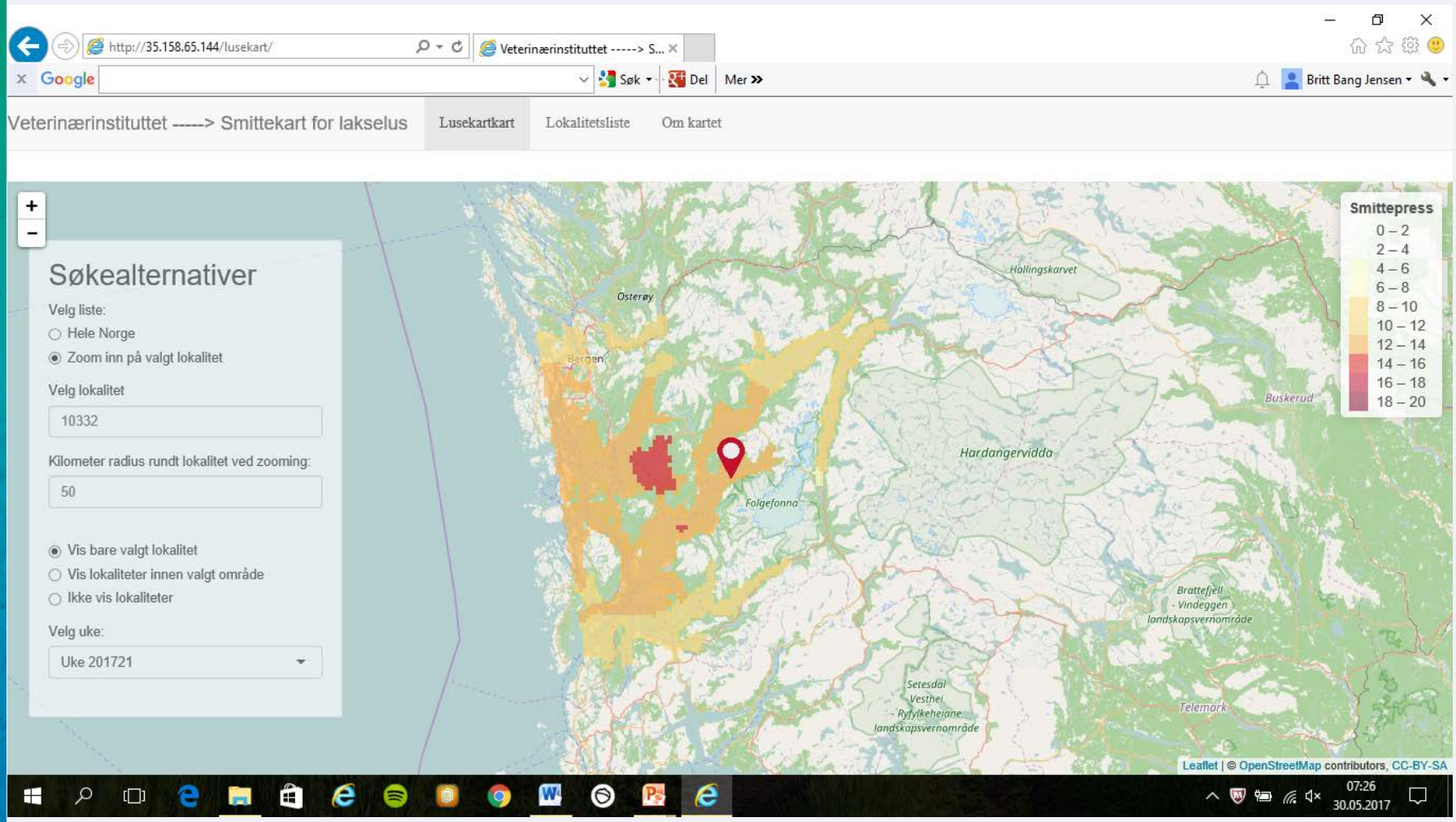






Implementation and dissimination: App with map of simulated salmon lice

- www.vetinst.no/lusekart



Norwegian Veterinary Institute

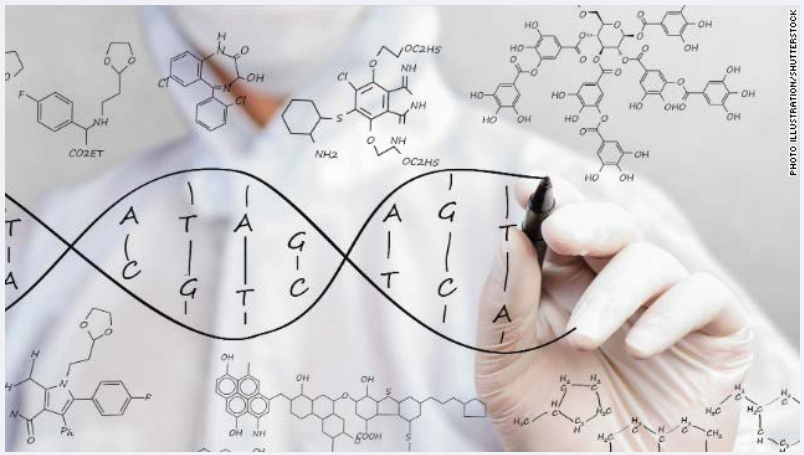
How to collate good data for outbreak investigation

- Test if the genotyping supports associations between virus outbreaks and risk factors for transmission

- Identify possible risk factors:
  - Smolt/egg supplier
  - Ownership
  - Broodstock
  - Geographical (Climate, river etc)

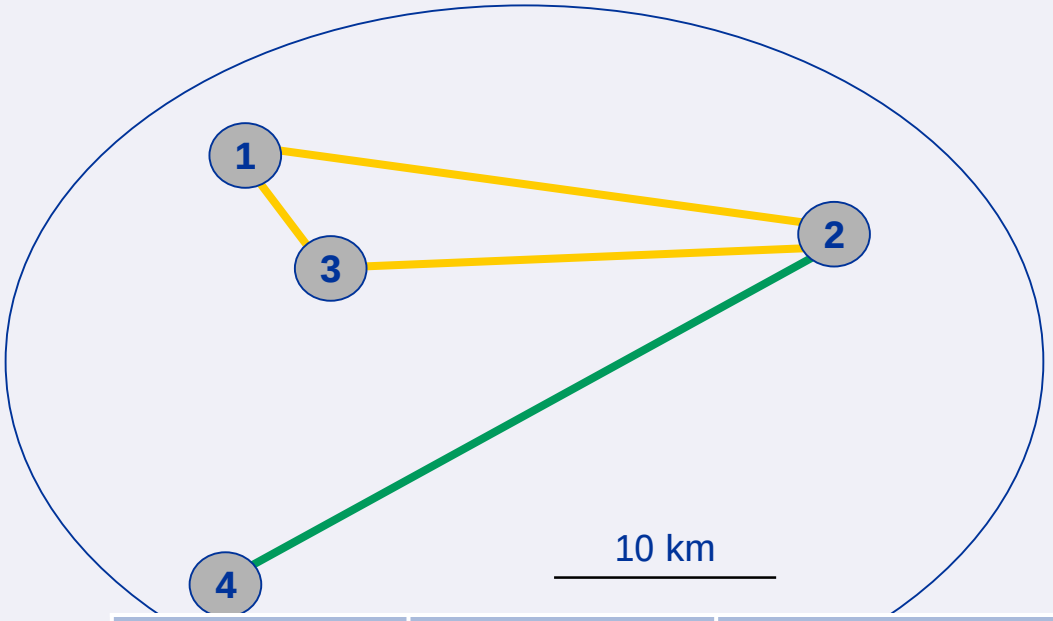
Outbreak investigation

- Assumptions:
  - any pair of virus isolates from different sites may or may not share a common source of infection…
  - directly sharing an infected source → genetically similar
  - not sharing an infected source → no expectations



Norwegian Veterinary Institute

Example of how data can be compiled



Outbreaks within the same 12 months

Outbreaks within the same concession (possible contact)

Outbreaks sharing smolt supplier

Matched outbreak	Proximity	Contact (Same concession)	Smolt supplier
1 & 2		x	
1 & 3	x	x	
1 & 4			
2 & 3		X	
2 & 4			x
3 & 4			

Norwegian Veterinary Institute

Statistical test

A binomial test can be used to test if the number of successes are higher than expected by chance for each of the three risk factors

S=success

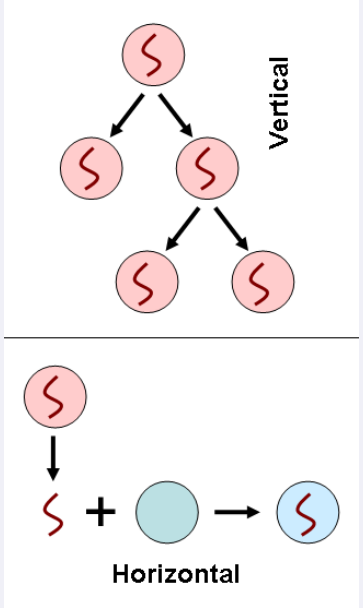
F=failure

Matched Outbreaks	Months apart	Genogroup	Proximity	Contact	Smolt Supplier
1 - 10	9	G3, G3			S
2 - 14	11	G1, G1	S		
3 - 7	1	G1, G1	S	S	
4 - 5	<1	G2, G2	S	S	
4 - 6	1	G2, G2			S
5 - 6	1	G2, G2	S	S	S
6 - 8	3	G2, G2	S	S	S
8 - 9	<1	G2, G3			F
10 - 16	3	G3, G3			S
11 - 15	3	G1, G3			F
12 - 20	9	G2, G3			F
13 - 18	8	G1, G1	S		
16 - 17	1	G3, G3			S
17 - 20	7	G3, G3			S
18 - 19	4	G1&G3, G1	S	S	
19 - 22	5	G1, G2			F
20 - 21	3	G3, G1			F
Summary statistics binomial test					
N			7	5	12
Observed S			7	5	7
Expected S			2.8	2.0	4.8
P			0.002	0.011	0.16

Norwegian Veterinary Institute

How to collate good data for molecular tracing

- Disease transmission:
  - Use the relationship between virus gene sequences to test different transmission pathways of virus
- Transmission pathways
  - Horizontal transmission
  - Vertical transmission
- Data required:
  - Cases
    - Incl. sequences
  - Population at risk



Challenges when collating data

- Only information on cases:
  - No knowledge of «population at risk»
  - You can test if cases are genetically similar
  - But not associations with risk factors
  - Tracing outbreaks is challenging:
    - For example: Similarity between genes from cases and smolt supplier does not necessarily mean that the virus came from this supplier if you have no knowledge of the status of other suppliers
  - Can help suggest possible risk factors

Challenges when collating data

- Problems with case definition:
  - Not notifiable disease
    - You do not know if all cases are reported -> controls can actually be cases
  - Clinical disease required or not?
  - Different practices at different labs:
    - Protocols/cut-off values/histopath definitions etc
  - Controls can turn into cases

Challenges when collating data

- Problems with production data:
  - No national/regional register
    - Required by EU-legislation
  - No knowledge on which farms are active
  - Information on production confidential to farmers
    - Can be solved though anonymization
  - Can be solved through cooperation with farmers and use of questionnaires
  - Maybe select a sub-set of the population at risk

Take home message:

- Quality of data is very important!
  - You must «know your data» – strengths and weaknesses
- Data must be fit for purpose
  - Formulate hypothesis/what do you want to know?
  - Then collect data
- But if you can't do that, see what you can get –often you can find something suitable….