This use case is about systems biology for salmon farming.

The Digital Salmon will be a knowledge base of salmon physiology, in the form of mathematical models linked to omics data.

It will help the salmon farming industry to navigate conflicting demands -- of sustainability, shifting feed prices, diseases and product quality. The industry needs to develop a flexible, integrated basis of knowledge for rapid response to new challenges.

And we will begin with the challenges of novel, sustainable feedstuffs.

Photo: Shutterstock
This work is currently funded by the Research Council of Norway, as part of their systems biology initiative called "Digital Life".
Participants at FAIRDOM User Meeting, 2016-09-15

Tina Graceline
Temporary dry-lab PAL
(project area liaison)
Bioinformatician

Jacob Torgersen
Wet-lab PAL
Molecular biologist
at AquaGen breeding company

Fabian Grammes
Leads "omics and imaging"
work package
Molecular biologist
and bioinformatician

Jon Olav Vik
Project leader, modeller
Ecologist turned systems biologist

This shows the project members who have been central in the generation, analysis and stewardship of data and models in DigiSal.
I will talk briefly about the importance of salmon farming and why it needs systems biology. Then I will outline what kinds of assets we have: our experiments, measurements and models. Finally, I'll share some challenges we have encountered in our use of the SEEK.
Salmon farming is big in Norway. One thousand fish farms employ 6000 people, who…

Photo: Shutterstock
...produce lots of salmon...

Photo: Shutterstock
…that become lots of dinners all over the world. The value generation from salmon farming is huge…

Photo: Shutterstock
…estimated at six thousand million euros per year.

Photo: Shutterstock
Salmon don't eat what they used to

Novel feedstuffs call for a systems biology approach

Now...
Salmon are carnivores by nature...

Photo: Shutterstock (salmon, sand eels), Wikipedia (herring, prawn, Gammarus)
… So in the early years of salmon farming, they were fed something… *similar* to their natural diet…

Phytoplankton
Primary producers of omega-3
Photo: Shutterstock (salmon, sand eels), Wikipedia (herring, prawn, Gammarus)
…until about 15 years ago.

Photo: Shutterstock
But now fish oil is a scarce resource. So instead, we taught salmon to eat…
…this. And today, about 75% of the fat and protein…

Photo: Shutterstock
…comes from plants. But it's not straightforward. Over a period of ten years, salmon feed has changed from containing ten ingredients with one protein source and one fat source, to more than thirty ingredients with several protein and fat sources.

Feedstuff prices fluctuate rapidly, and it is time-consuming and expensive to trial new diets on fish.

Also, even plant-fed salmon is not particularly sustainable, because salmon now competes with plant production for human food.

Therefore, people are now exploring novel, sustainable feed ingredients…

Photo: Shutterstock
Novel, sustainable feedstuffs

Yeast • Bacterial meal • Microalgae

Sustainable feed ingredients from natural bioresources that are not suitable for direct human consumption

http://www.foodsofnorway.net

...such as yeast, bacterial meal and microalgae.

But the effects of the novel feed ingredients on the salmon body are complex and involve many organs.
And this brings us to systems biology, understanding the living body as a set of components that both affect each other and depend on each other.

Systems biology is about identifying components, the processes through which they interact, identifying their interfaces, and model how organs, tissues, cells and organelles respond to stimuli.

This involves a lot of data of different kinds, and needs modelling of many different processes. Obviously, our data and model management relies critically on findable, accessible, interoperable and reusable data and models.

We chose to focus first on metabolic reaction networks, since that is a phenomenon and a process that we can manipulate, measure and model.

At the moment, we specifically look at the synthesis of healthy omega-3 fatty acids.
What we manipulate, measure and model
The old school feeding trial means that you feed different diets to fish, wait weeks or months, then measure things like body weight and nutritional value of the fish at the end.

As a faster, cheaper alternative, Jacob also grows slices of liver in the lab, keeping it alive and feeding it in vitro. This allows us to try many diets on the same individual fish. We can measure metabolic activity by sacrificing some slices for RNA sequencing and metabolomics.

Here's what it looks like. You cut out a small block of liver and superglue it to a plastic cylinder. Then you encase it in agar and mount it in a vibratome, a slicing machine. Each slice goes into its own lab dish, to be fed for example vegetable oil or marine oils.

The iteration between experiment and modelling will be much faster than for the feeding trials, informing those longer-running, more expensive trials to increase their usefulness.

Photo: Jon Olav Vik
Measurements of gene expression and metabolite concentrations enable us to mathematically model the biochemical reaction network. Gene expression tells us which reactions are active in a given tissue on a given diet, and metabolomics tells us how much we have of the various metabolites. We can then compute what are the possible flux patterns through the reaction network that are consistent with our observations. A little data will give rough constraints on what is happening, more data will give a clearer picture.

The analysis of these models will suggest hypotheses and designs for new experiments.
So: DigiSal is a **model-driven, tightly integrated theoretical-experimental study of mechanistic interactions among genetic and feed factors**. By combining experiment and modelling we aim to deliver a predictive understanding of a whole range of possible diets, much more efficiently than by traditional feeding trials alone.

From a **FAIRDOM perspective**, we have here a lot of **data and models that can potentially be linked via common languages**: Genes code for enzymes, which catalyze biochemical reactions, which transform molecules whose concentrations we can manipulate, measure and model.

Tina, as our dry-lab PAL, helps researchers get their data into the SEEK in standard formats, together with Fabian and Jacob, who lead the work packages on omics and tissue-slice culture.

As you see, this is a rich and challenging project, where modelling is the driving engine, and FAIRDOM is a key enabling infrastructure.

Gareth mentioned the importance of dedicating staff to data and model management.

We have indeed allocated a full-time position to this. However, for Tina this is slightly beside her chosen career path. And so **we are hiring**. We need someone to champion FAIRDOM in this position [point], as well as two modellers to drive the research cycles [point].
This is a timeline of our interaction with FAIRDOM.
Our Digital Salmon vision was presented at a symposium in early 2013 and subsequently published as a book chapter.

I first encountered FAIRDOM when we wrote our SysFat application to the first ERASysAPP call in January 2014. That call required all proposals to formulate a data and model management strategy, and to submit to the guidance and supervision of the central data management support group. This was the first time I'd seen a call that encouraged *allocating resources* to data and model management; where you could be proud of spending the extra time required to ensure reusability and proper documentation.

Now, although SysFat was not funded, we did not throw away this data and model management plan. We used it when we reworked our proposal into "GenoSysFat", which was submitted in fall 2014 and funded in December. We also used the same strategy when developing our Digital Life proposal in 2014. During this time, we attended a workshop in Rostock, where we asked so many questions that we were eventually invited to become a FAIRDOM Partner. DigiSal got funded, and Natalie and Stuart took part in our kick-off meeting and presented FAIRDOM and the SEEK.

And at the first FAIRDOM User Meeting in Barcelona, I presented the Digital Salmon as a Star Use Case.
Support from FAIRDOM

Natalie Stanford and Stuart Owen are our contacts in FAIRDOM, providing training and advice and forwarding feature requests.

Set up projects GenoSysFat and DigiSal at fairdomhub.org.

Both attended the DigiSal kick-off meeting in March 2016, where Natalie gave a presentation on knowledge management and FAIRDOM.

Helped adapt the ISA (Investigation-Study-Assay) model to our use case.
Norwegian followup

Since the inception of Digital Life Norway, DigiSal has advocated a strategy for interoperability and knowledge management also vs the Research Council of Norway

Some of the new Digital Life projects are planning to adopt FAIRDOM

Dialogue between FAIRDOM, Digital Life Norway, and the Research Council of Norway

Integration with the Norwegian e-infrastructure for Life Sciences, NeLS
Data and model management in the Digital Salmon
The Digital Salmon is currently made up by two sister projects with overlapping teams and similar goals. As you see, we have begun to accumulate a number of data files, organized in an Investigation - Study - Assay structure.
We had a good dialogue with Natalie on how to map the ISA structure onto our own needs. This is our interpretation, and we're reasonably happy with it.

Adapted from ISA best practice and programmes and projects
What I really like about the SEEK is how it clarifies what we’re talking about. No more "can we discuss ‘the gene expression data’? [airquotes]. Instead I just copy the hyperlinked text [point] into an email. The ISA graph places every asset in context, saving us from lots of explanations and misunderstandings.

(Since this screenshot, the system has been upgraded to show more levels simultaneously, including Data Files, SOPs and Projects.)

What the screenshot does not show is that the Data Files also tie to external databases, because they use standardized identifiers for genes and so on.
Looking forward to future developments

Programmatic access to datasets in SEEK e.g. from R

How to handle **big, non-public assets** that reside e.g. on our own cluster
(ongoing dialogue between FAIRDOM and Norwegian e-infrastructure for Life Sciences, NeLS)

How to handle **multifile assets**, e.g. RNAseq data shoehorned into MIAME format
Our ongoing quest for a knowledge manager (biocurator, FAIRDOM evangelist)

A full time position that we’re struggling to fill.

Announced Oct 2015, April 2016 with advice from EBI, Leiden, FAIRDOM, ...
Seeking PhD or MSc, biosemantics, bioinformatics or similar
- Work with researchers to clarify their data management needs and relevant standards
- Use semantic queries and automated reasoning to reap the benefits of ontological annotation and semantic interoperability

Very hard to find people with the right skillset/interests, or to be noticed by them (NMBU, Norway may seem a bit remote)

The position as biocurator and FAIRDOM evangelist plays two crucial roles: One is to work with researchers to clarify their data management needs and relevant standards. The other is using semantic queries and reasoning to reap the benefits of ontological annotation and semantic interoperability.

Also, we need two systems biologists to analyse the exciting data we acquire. One focus is metabolic modelling; others include exploratory multivariate analysis as well as evolutionary genomics.

So: Do you know a modeller or biosemantician who might be interested? Please spread the word!

Thank you. Now, back to the talk...
Many people and institutions are involved in this project, and we welcome input from other interested parties. Visit our web page and have a look! Thank you.

Photo: Shutterstock
Lightbox (picture collection) at Shutterstock:
http://www.shutterstock.com/public_lightbox.mhtml?lightbox_id=43925734&code=a274444ce9392ea48afbb2e2b2e6c5c6