

The challenge of sensor selection, long term-sensor operation and data evaluation in inter- -institutional long term monitoring projects

- lessons learned in the MOSES project -



Philipp Fischer¹, Madlen Friedrich¹, Markus Brand¹, Uta Ködel², Peter Dietrich², Holger Brix³, Dorit Kerschke⁴, Ingeborg Bussmann¹ and the MOSES & Digital Earth Team

¹ Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung (AWI)

² Helmholtz-Zentrum für Umweltforschung (UFZ)

³ Helmholtz-Zentrum Geesthacht, Institute for Coastal Research (HZG)

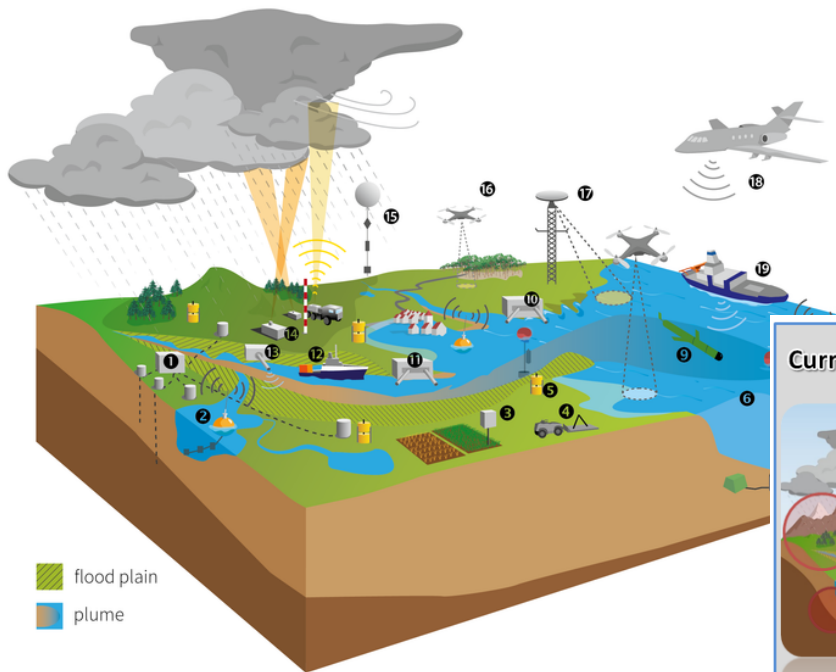
⁴ Helmholtz-Zentrum Potsdam (GFZ)

- Today, **Sciences Across Disciplines**...is a „buzz“-term in many national and international science programs.
- The motivation of such “across disciplines” projects is the scientific awareness that functions and processes in most marine and terrestrial ecosystems are highly complex and interact across our human definitions of scientific disciplines.

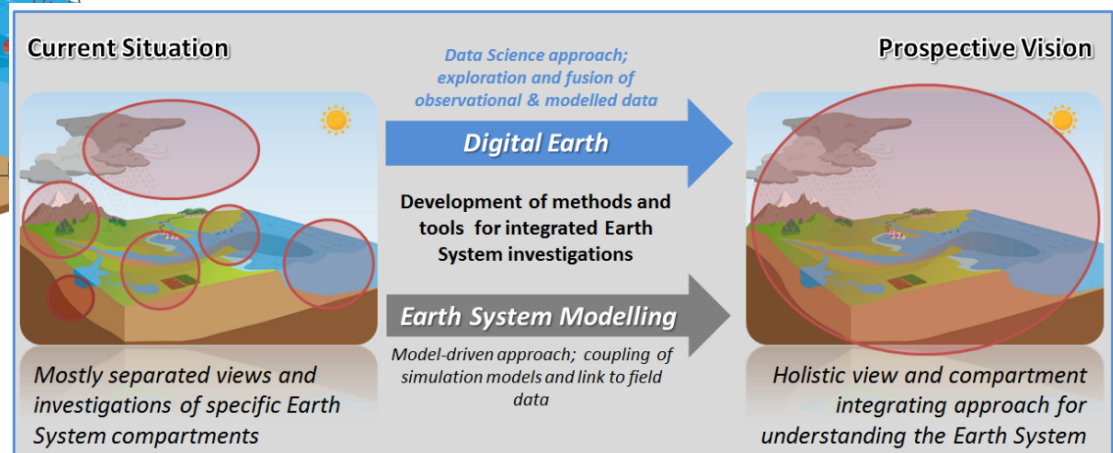
Cross discipline approaches are “en vogue”



MOSES and Digital Earth are two novel observing „systems“ and „strategies“ of the German Helmholtz Association specifically designed to investigate the interactions of short-term events and long-term trends **across Earth compartments.**



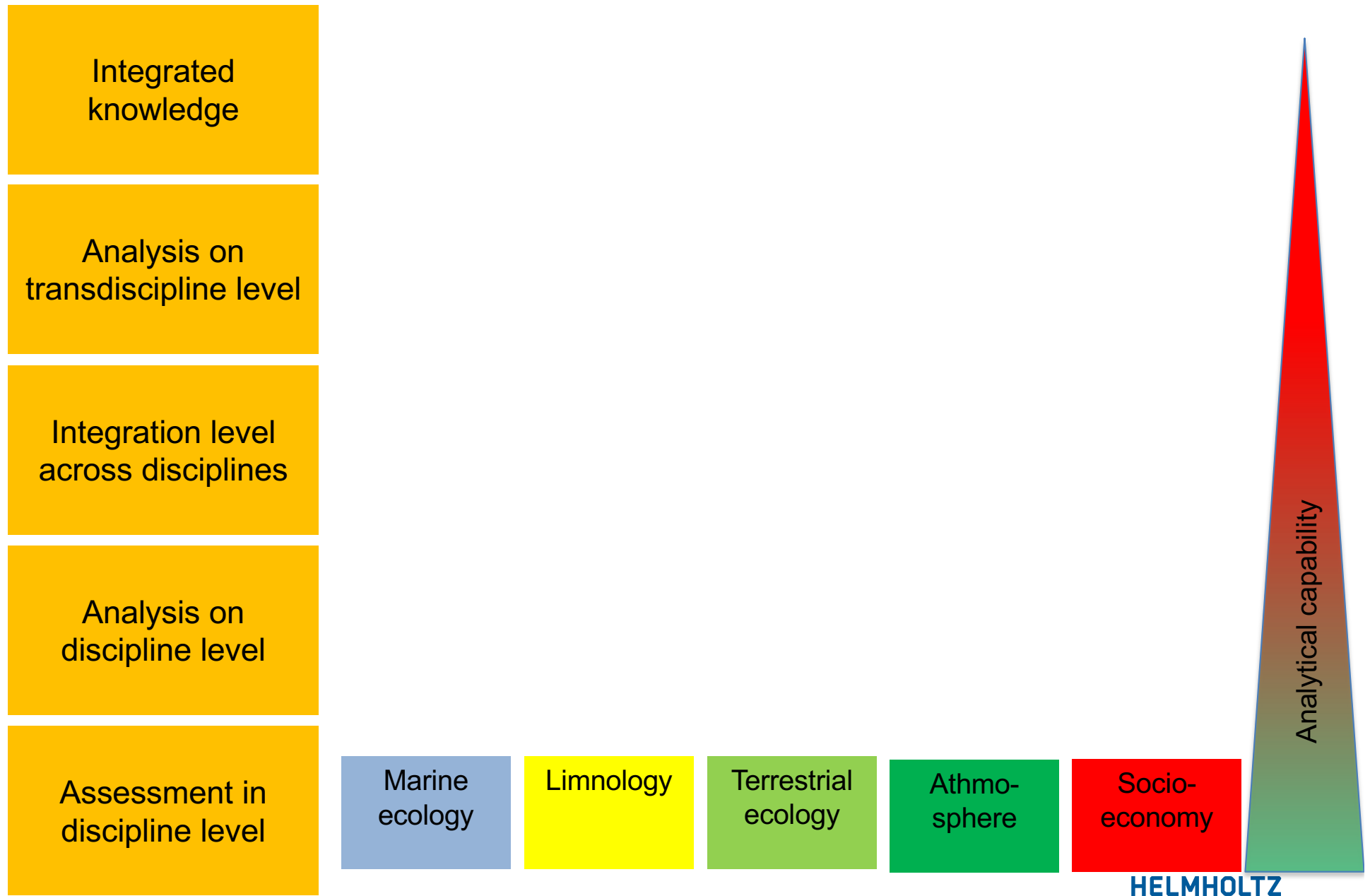
■ flood plain
■ plume



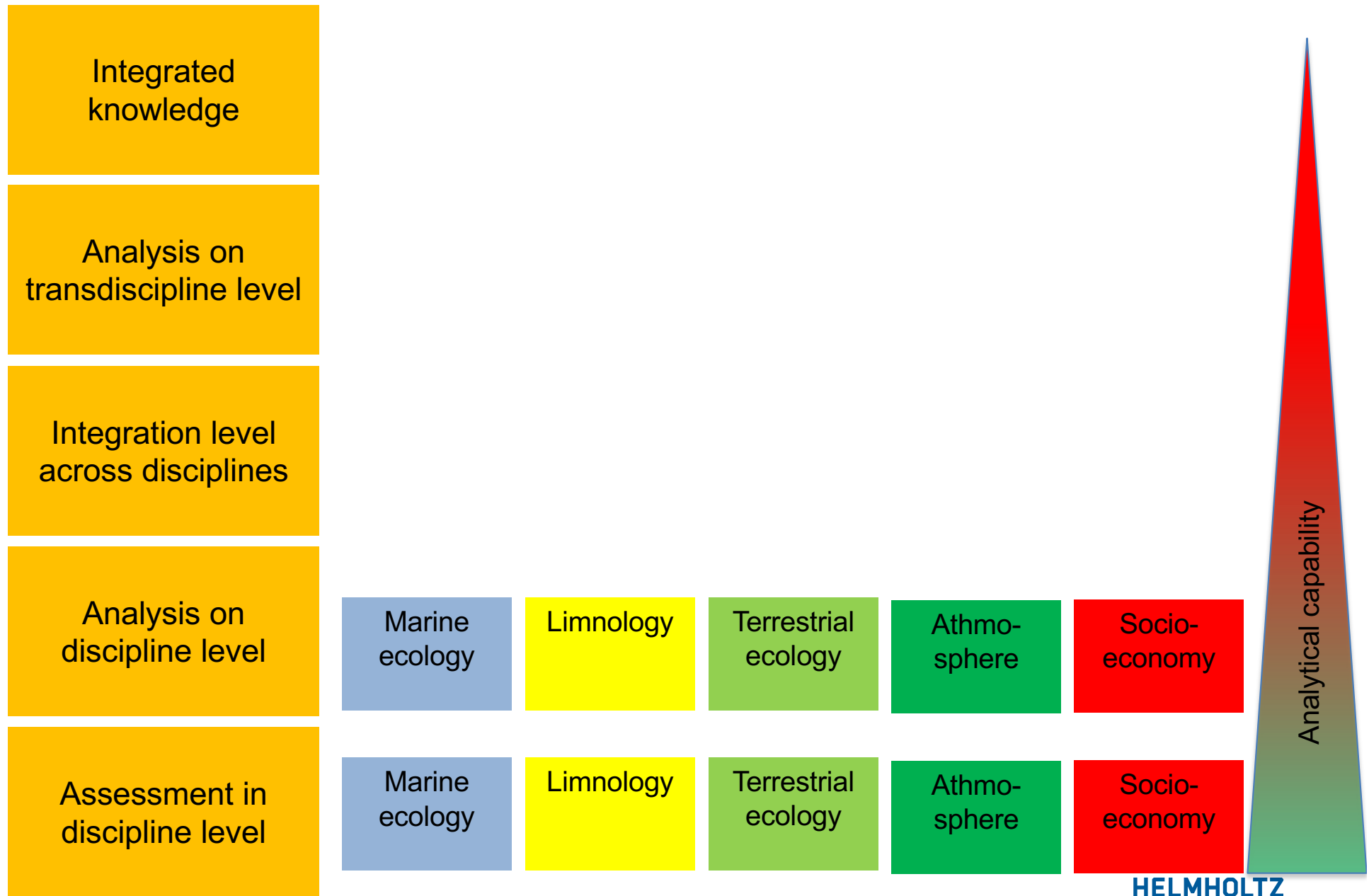
The term “across disciplines” itself however implies that we go beyond our scientific knowledge.... and it is not always very clear...

how to do this?

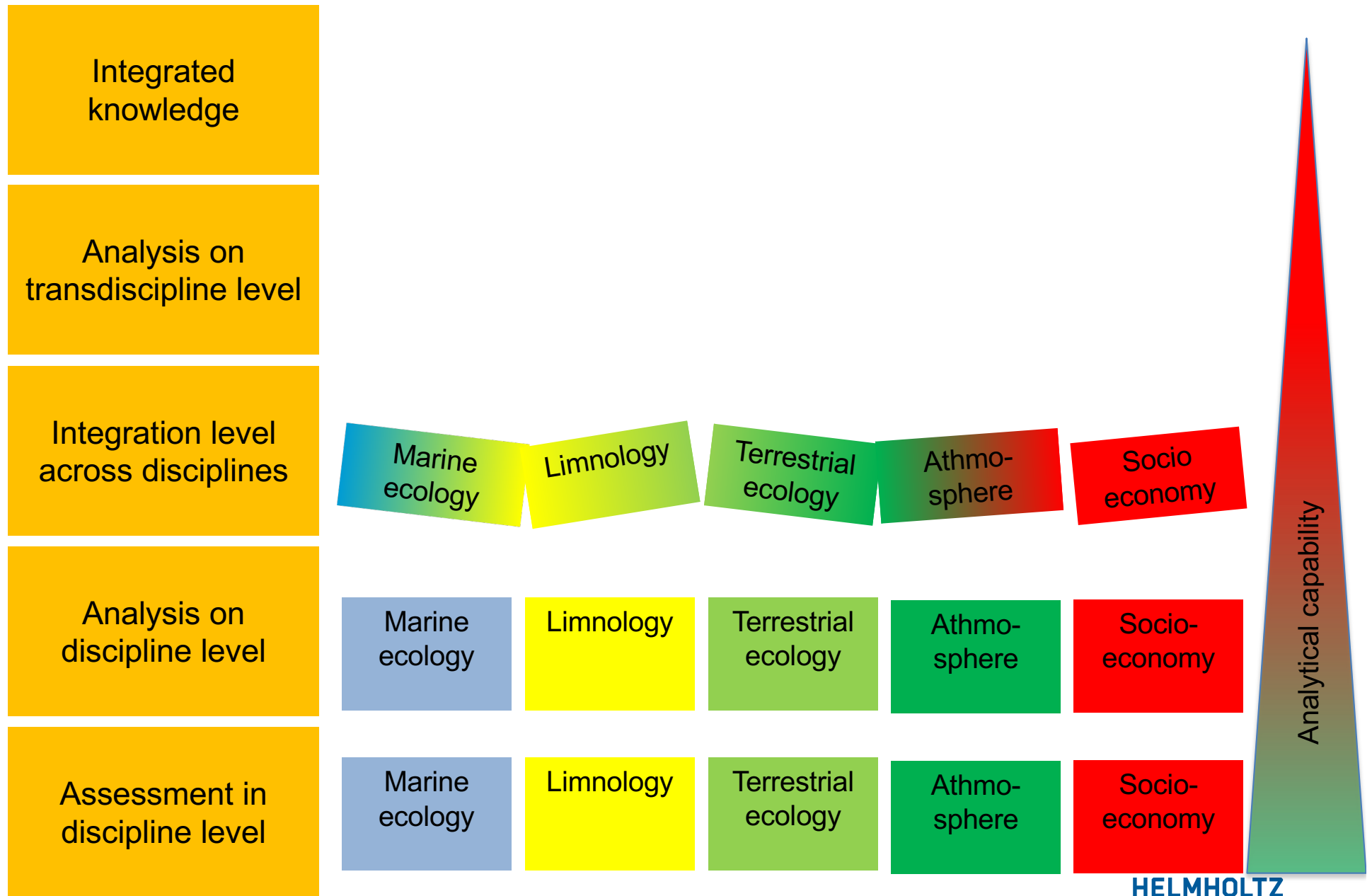
Ideal workflow



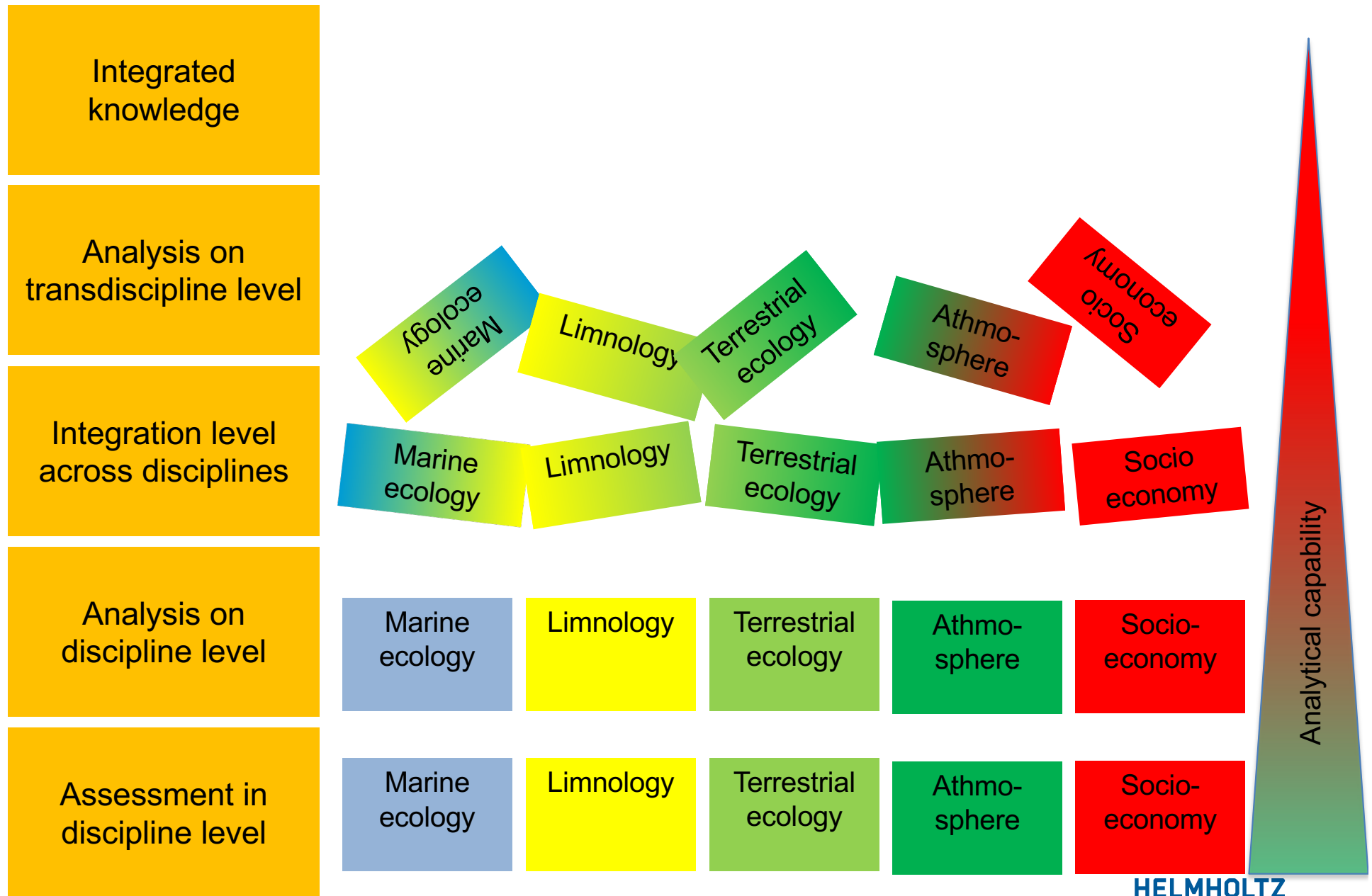
Ideal workflow



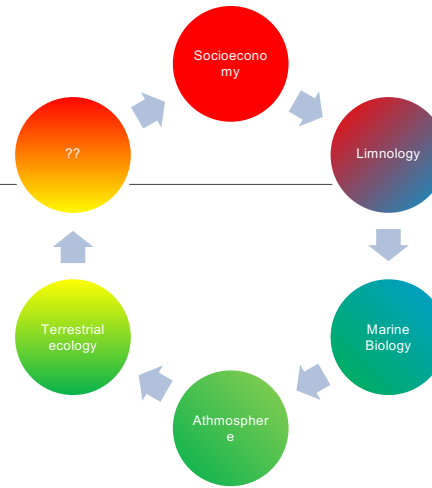
Ideal workflow



Ideal workflow



Ideal workflow



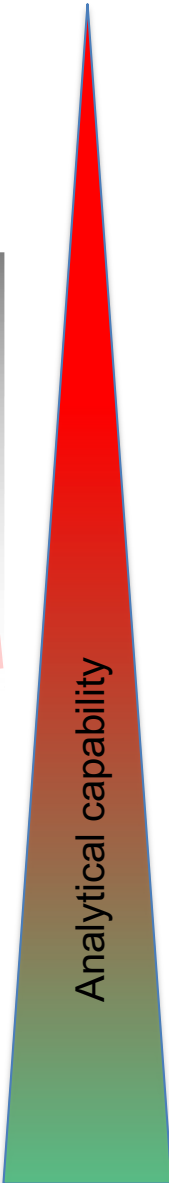
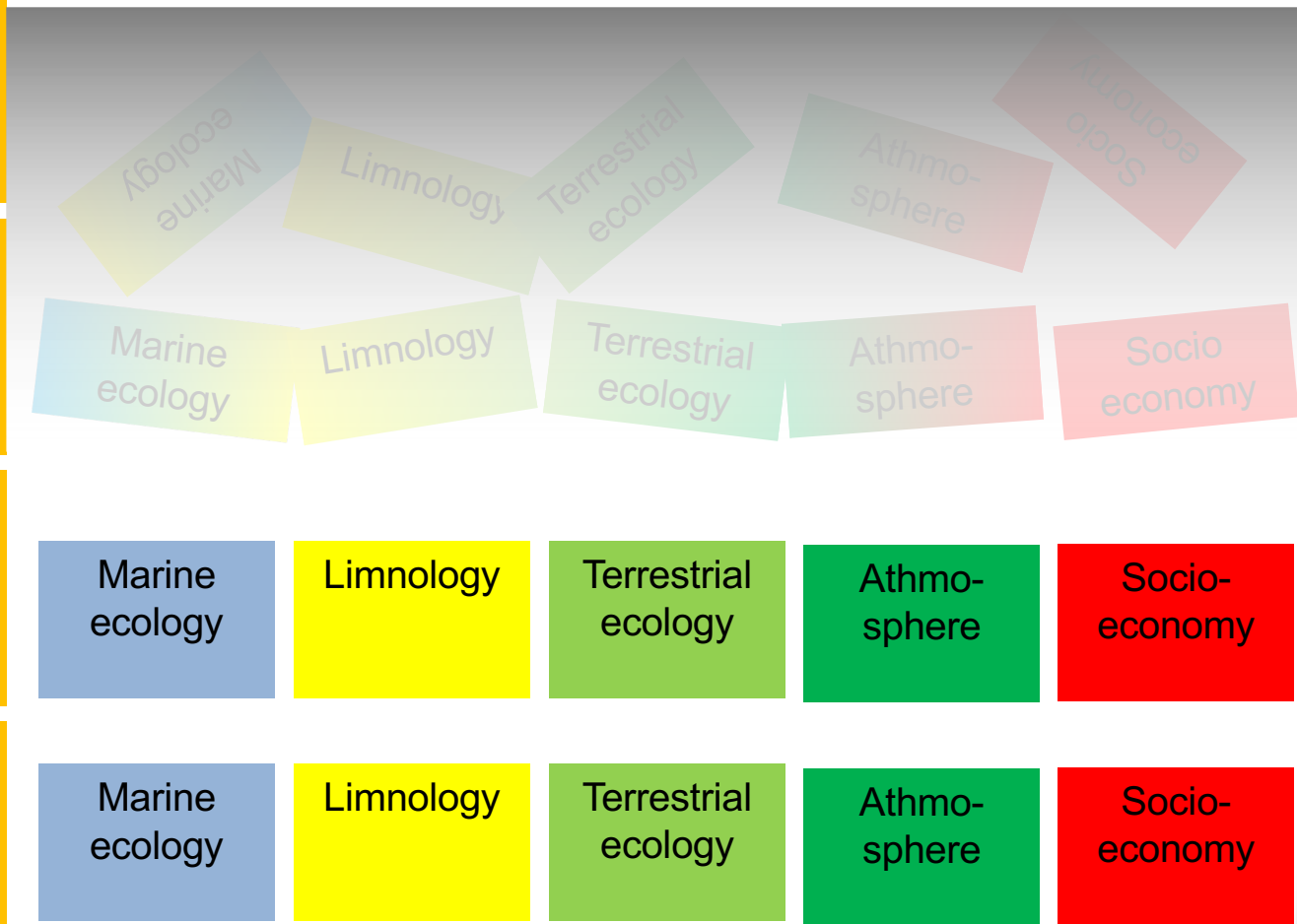
Integrated knowledge

Analysis on transdiscipline level

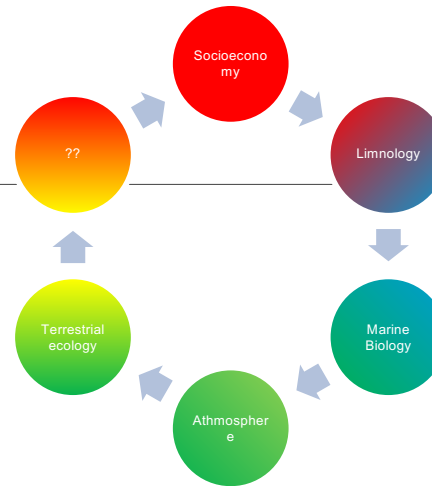
Integration level across disciplines

Analysis on discipline level

Assessment in discipline level



Ideal workflow



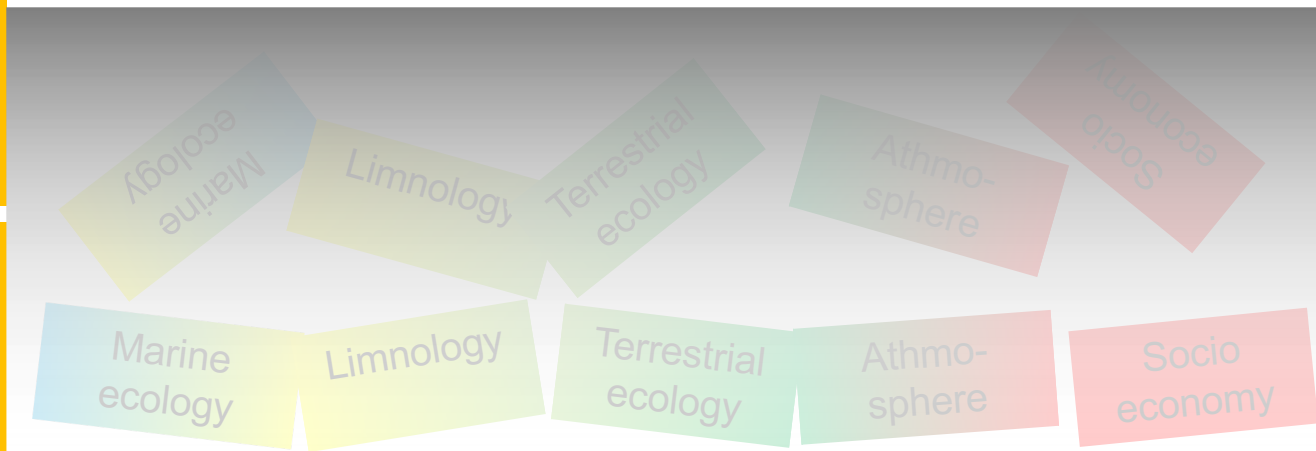
Integrated knowledge

Analysis on transdiscipline level

Integration level across disciplines

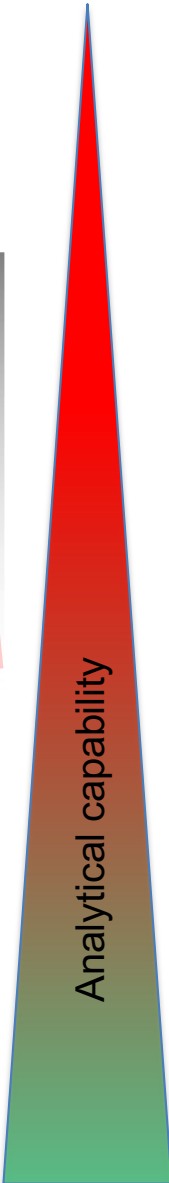
Analysis on discipline level

Assessment in discipline level

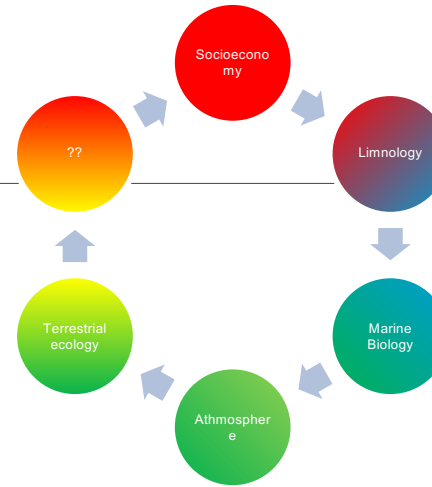


Our comfort zone

Marine ecology	Limnology	Terrestrial ecology	Athmo-sphere	Socio-economy
Marine ecology	Limnology	Terrestrial ecology	Athmo-sphere	Socio-economy



Ideal workflow



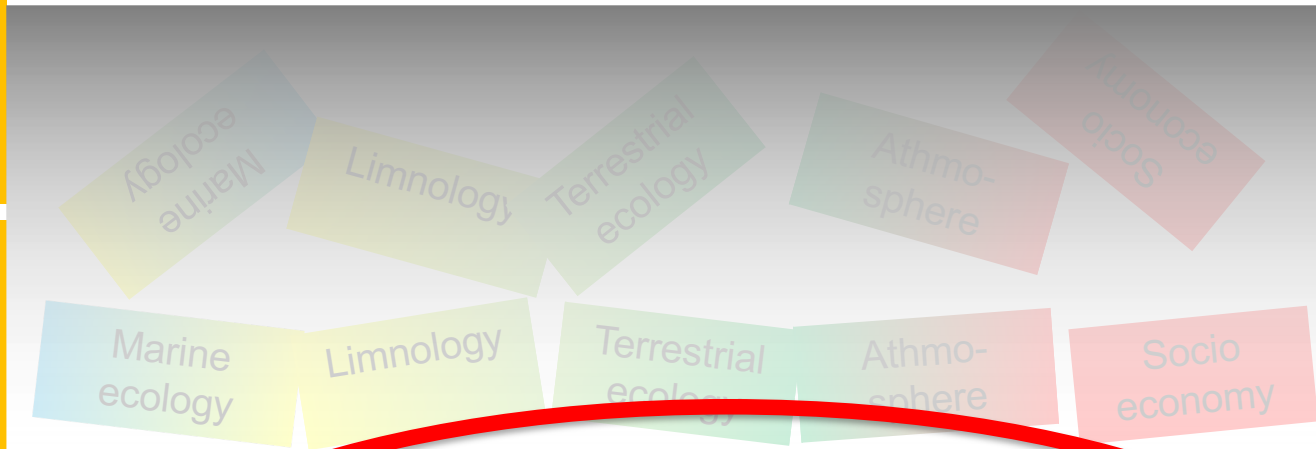
Integrated knowledge

Analysis on transdiscipline level

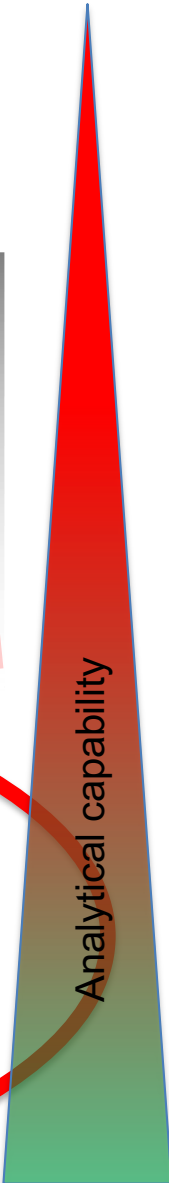
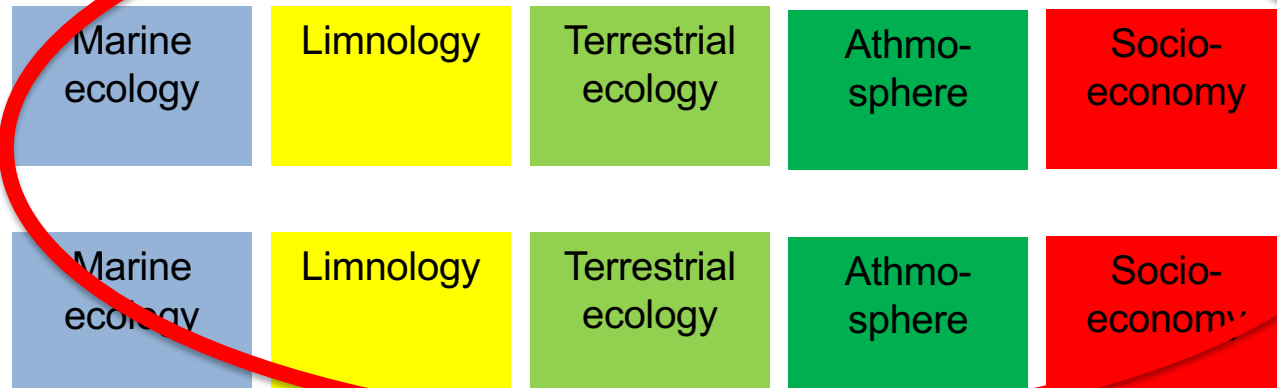
Integration level across disciplines

Analysis on discipline level

Assessment in discipline level



Our comfort zone



Cross-cutting problems



- Independent of the scientific discipline, the same problems emerge in almost any discussion.
 - Data processing with respect to data quality.
 - Stable data assessment in (near-realtime).
 - Integration, aggregation & analysis of simple and complex data.

Data processing with respect to data quality....

In July 2018 and November 2019, two joint workshops of the AWI, HZG, GEOMAR, UFZ and GFZ were organized.

The main goals of these workshops were:

- to test the comparability of different sensors measuring the same parameters (e.g., water temperature, oxygen, etc.)
- to test the MOSES data flows from the sensors to the database with respect to an inter-institutional availability of the obtained sensor data,
- to test new MOSES sensors under controlled conditions and to development a MOSES concept (Standard Operational Procedures - SOPs) for the intercalibration of sensors before joint measurement campaigns.

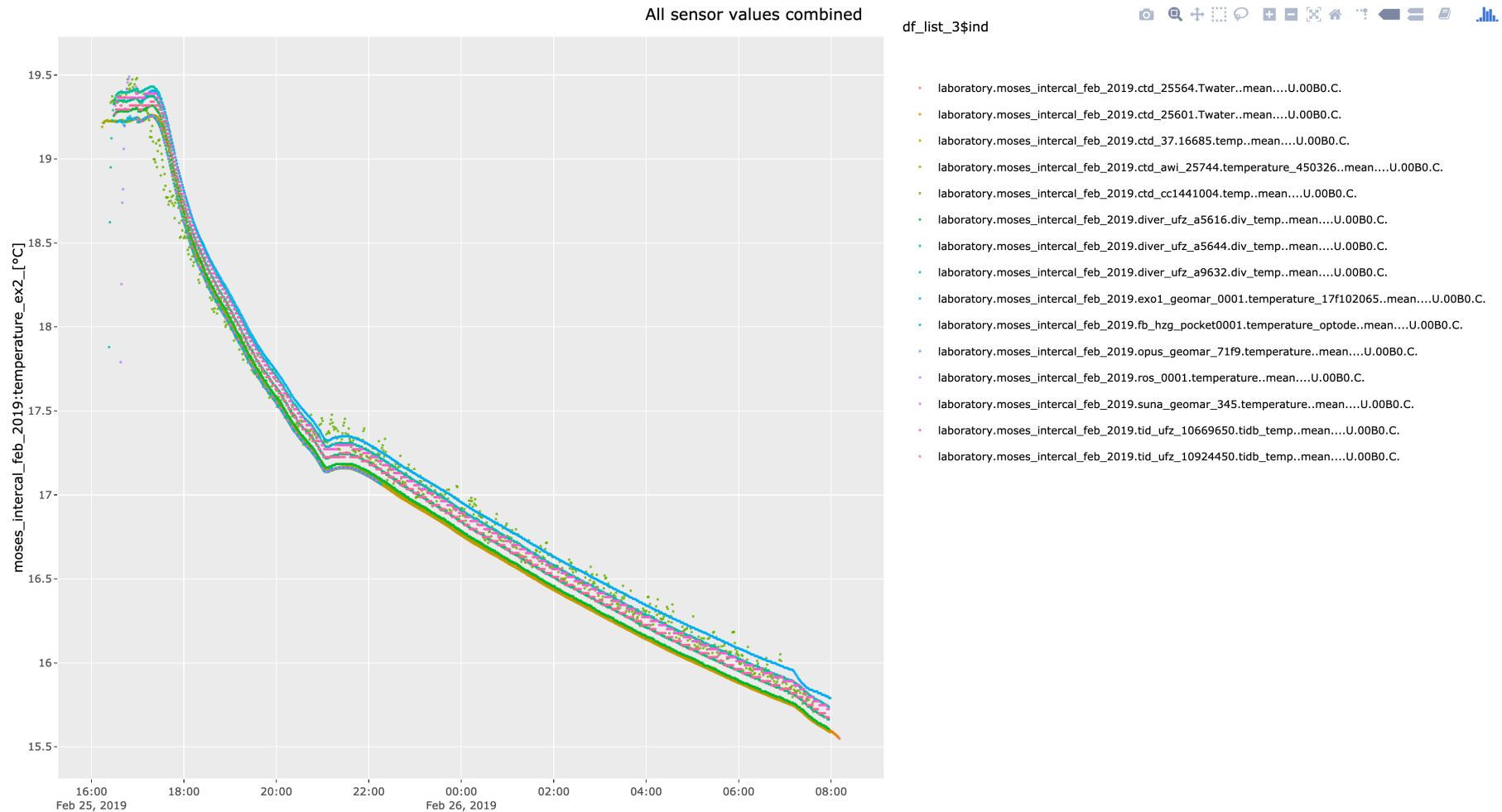
A total of 15 sensors with different parameters were tested in a simultaneous approach.

- AWI • laboratory.moses_intercal_feb_2019.ctd_25564.Twater..mean....U.00B0.C.
- AWI • laboratory.moses_intercal_feb_2019.ctd_25601.Twater..mean....U.00B0.C.
- AWI • laboratory.moses_intercal_feb_2019.ctd_37.16685.temp..mean....U.00B0.C.
- AWI • laboratory.moses_intercal_feb_2019.ctd_awi_25744.temperature_450326..mean....U.00B0.C.
- HZG • laboratory.moses_intercal_feb_2019.ctd_cc1441004.temp..mean....U.00B0.C.
- UFZ • laboratory.moses_intercal_feb_2019.diver_ufz_a5616.div_temp..mean....U.00B0.C.
- UFZ • laboratory.moses_intercal_feb_2019.diver_ufz_a5644.div_temp..mean....U.00B0.C.
- UFZ • laboratory.moses_intercal_feb_2019.diver_ufz_a9632.div_temp..mean....U.00B0.C.
- GEOMAR • laboratory.moses_intercal_feb_2019.exo1_geomar_0001.temperature_17f102065..mean....U.00B0.C.
- HZG • laboratory.moses_intercal_feb_2019.fb_hzg_pocket0001.temperature_optode..mean....U.00B0.C.
- GEOMAR • laboratory.moses_intercal_feb_2019.opus_geomar_71f9.temperature..mean....U.00B0.C.
- ROSTOCK • laboratory.moses_intercal_feb_2019.ros_0001.temperature..mean....U.00B0.C.
- GEOMAR • laboratory.moses_intercal_feb_2019.suna_geomar_345.temperature..mean....U.00B0.C.
- UFZ • laboratory.moses_intercal_feb_2019.tid_ufz_10669650.tidb_temp..mean....U.00B0.C.
- UFZ • laboratory.moses_intercal_feb_2019.tid_ufz_10924450.tidb_temp..mean....U.00B0.C.

n = 15

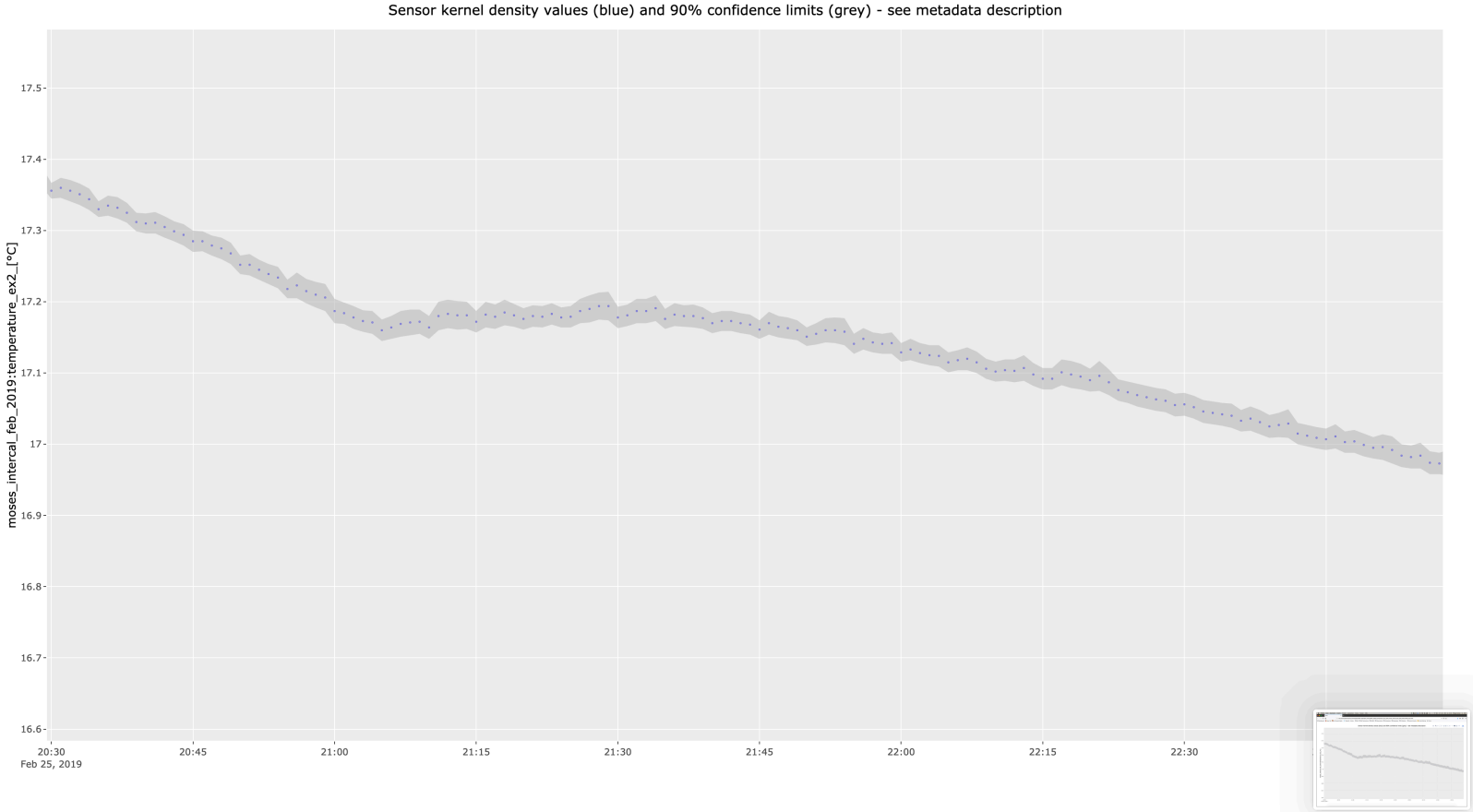
Data processing with respect to data quality....

Focussing on temperature:



Data processing with respect to data quality....

Intelligent data compilation by statistical procedures like kernel density estimates and / or auxiliary sensor use allow for modelling / predicting missing data, sensor drift **AND most important Accuracy and Precision information.**



Data processing with respect to sensor plausability...



Applying ARGO plausability checks:

par_id	All checks passed	Failed at quality check level 11 (Gradient test)	Failed at quality check level 7 (Regional range test)	Total
laboratory.moses_intercal_feb_2019.ctd_25564.Twater..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.ctd_25601.Twater..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.ctd_37.16685.temp..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.ctd_awi_25744.temperature_450326..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.ctd_cc1441004.temp..mean....U.00B0.C.	99.78190	0.1090513	0.1090513	100
laboratory.moses_intercal_feb_2019.diver_ufz_a5616.div_temp..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.diver_ufz_a5644.div_temp..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.diver_ufz_a9632.div_temp..mean....U.00B0.C.	99.78678	0.1066098	0.1066098	100
laboratory.moses_intercal_feb_2019.exo1_geomar_0001.temperature_17f102065..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.fb_hzg_pocket0001.temperature_optode..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.opus_geomar_71f9.temperature..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.ros_0001.temperature..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.sun_a_geomar_345.temperature..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.tid_ufz_10669650.tidb_temp..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
laboratory.moses_intercal_feb_2019.tid_ufz_10924450.tidb_temp..mean....U.00B0.C.	100.00000	0.0000000	0.0000000	100
Total	99.96106	0.0194723	0.0194723	100

The real data

Data processing with respect to sensor plausability...

Additional statistical procedures...

laboratory:moses_intercal_feb_2019

Sensor URN	Adjusted R-squared	Statistical significance of that regression between measured and expected value - variables are correlated	Slope of regression of measured values against expected value	Statistical significance of difference in slope of measured values and expected values	Median accuracy (difference between measured value and expected value = median residuen)	Mean accuracy (difference between measured value and expected value = mean residuals)	Statistical significance of accuracy (mean difference between measured values and expected value)	Precision (+/- spread around measured value (95pct confidence limit))	Coefficient of variation (pct) of precision to accuracy
moses_intercal_feb_2019:ctd_awi_25744:temperature_450326	0.9996052	***	1.0053909	***	-0.0009073	-0.0091557	***	0.0015644	17.0867988
moses_intercal_feb_2019:diver_ufz_a9632:div_temp	0.9976658	***	1.0080419	***	0.0016729	0.1140364	***	0.0032114	2.8161484
moses_intercal_feb_2019:ctd_cc1441004:temp	0.9926212	***	0.9562691	***	-0.0020758	0.0767775	***	0.0061902	8.0624610
moses_intercal_feb_2019:diver_ufz_a5616:div_temp	0.9999524	***	1.0165609	***	0.0005371	-0.0082756	***	0.0011634	14.0582718
moses_intercal_feb_2019:diver_ufz_a5644:div_temp	0.9999473	***	1.0169230	***	0.0003625	0.0469419	***	0.0011941	2.5437916
moses_intercal_feb_2019:tid_ufz_10924450:tidb_temp	0.9999064	***	1.0038544	***	-0.0003160	0.0452123	***	0.0006742	1.4912897
moses_intercal_feb_2019:ctd_25564:Twater	0.9999425	***	1.0028887	***	-0.0006834	-0.0244798	***	0.0005239	2.1399671
moses_intercal_feb_2019:ctd_25601:Twater	0.9999401	***	1.0032472	***	-0.0006723	-0.0244912	***	0.0005417	2.2118492
moses_intercal_feb_2019:tid_ufz_10669650:tidb_temp	0.9999089	***	1.0067264	***	-0.0000784	0.0991877	***	0.0007551	0.7612485
moses_intercal_feb_2019:exo1_geomar_0001:temperature_17f102065	0.9997965	***	0.9938749	***	-0.0002065	-0.0323417	***	0.0012814	3.9620727
moses_intercal_feb_2019:ros_0001:temperature	0.0524740	ns	-69.8181698	***	-0.0848775	-0.1951538	ns	0.3217956	164.8933121
moses_intercal_feb_2019:opus_geomar_71f9:temperature	0.9998785	***	0.9963838	*	0.0008304	-0.0207693	***	0.0021471	10.3380232
moses_intercal_feb_2019:ctd_37:16685	0.9999204	***	1.0010742	***	0.0001192	-0.0384071	***	0.0005576	1.4518944
moses_intercal_feb_2019:suna_geomar_345:temperature	0.9998968	***	0.9942930	***	-0.0000331	-0.0220024	***	0.0020931	9.5131915
moses_intercal_feb_2019:fb_hzg_pocket0001:temperature_optode	0.9998817	***	0.9921091	***	0.0011739	0.1568924	***	0.0008014	0.5107953

Data processing with respect to sensor plausability...

Additional statistical procedures and their possible application...

laborator

I need to determine the “real” temperature as accurate as possible and have less strict requirements for the variability of the measurements due to sensor specifications

Sensor URN	value	correlated	value	values	Median accuracy (difference between measured value and expected value = median residual)	Mean accuracy (difference between measured value and expected mean residuals)	Statistical significance (mean difference between measured values and expected values)	Precision +/- spread around measured value (95pct confidence limit))	Coefficient of variation (pct) of precision to accuracy
moses_intercal_feb_2019:ctd_awi_25744:temperature_450326	0.9996052	***	1.0053909	***	-0.0009073	-0.0091557	***	Unit price: 15 000 €	
moses_intercal_feb_2019:diver_ufz_a9632:div_temp	0.9976658	***	1.0080419	***	0.0016729	0.1140364	***	0.0032114	2.8161484
moses_intercal_feb_2019:ctd_cc1441004:temp	0.9926212	***	0.9562691	***	-0.0020758	0.0767775	***	0.0061902	8.0624610
moses_intercal_feb_2019:diver_ufz_a5616:div_temp	0.9999524	***	1.0165609	***	0.0005311	-0.0082756	***	Unit price: 250 €	
moses_intercal_feb_2019:diver_ufz_a5644:div_temp	0.9999473	***	1.0169230	***	0.0003625	0.0469419	***	0.0011941	2.5437916
moses_intercal_feb_2019:tid_ufz_10924450:tidb_temp	0.9999064	***	1.0038544	***	-0.0003160	0.0452123	***	0.0006742	1.4912897
moses_intercal_feb_2019:ctd_25564:Twater	0.9999425	***	1.0028887	***	-0.0006834	-0.0244798	***	0.0005239	2.1399671
moses_intercal_feb_2019:ctd_25601:Twater	0.9999401	***	1.0032472	***	-0.0006723	-0.0244912	***	0.0005417	2.2118492
moses_intercal_feb_2019:tid_ufz_10669650:tidb_temp	0.9999089	***	1.0067264	***	-0.0000784	0.0991877	***	0.0007551	0.7612485
moses_intercal_feb_2019:exo1_geomar_0001:temperature_17f102065	0.9997965	***	0.9938749	***	-0.0002065	-0.0323417	***	0.0012814	3.9620727
moses_intercal_feb_2019:ros_0001:temperature	0.0524740	ns	-69.8181698	***	-0.0848775	-0.1951538	ns	0.3217956	164.8933121
moses_intercal_feb_2019:opus_geomar_71f9:temperature	0.9998785	***	0.9963838	*	0.0008304	-0.0207693	***	0.0021471	10.3380232
moses_intercal_feb_2019:ctd_37:16685	0.9999204	***	1.0010742	***	0.0001192	-0.0384071	***	0.0005576	1.4518944
moses_intercal_feb_2019:suna_geomar_345:temperature	0.9998968	***	0.9942930	***	-0.0000331	-0.0220024	***	0.0020931	9.5131915
moses_intercal_feb_2019:fb_hzg_pocket0001:temperature_optode	0.9998817	***	0.9921091	***	0.0011739	0.1568924	***	0.0008014	0.5107953

Data processing with respect to sensor plausability...

Additional statistical procedures and their possible application...

laboratory:moses_intercal_feb_2019

I need to determine very small scale temperature changes over time without having the need of an accurate absolute temperature (climate change questions).

Sensor	Adjusted	Statistical	Statistical	Median	Mean	Statistical	Precision	Coefficient
			of significance	accuracy	accuracy	significance	(+/- spread	of variation
			of difference	(difference	(difference	of accuracy	around	(pct) of
			measured	between	between	(mean	measured	precision to
			in slope of	measured	measured	difference	value	(pct) of
			values and	value and	value and	between	(95pct	precision to
			inst values and	value =	value =	measured	confidence	accuracy
			ected expected	median	mean	values and	limit))	
			value values	residuen)	residuals)	expected		
						value)		
moses_			909 ***	-0.0009673	-0.0091557	***	0.0015644	17.0867988
moses_			419 ***	0.0016729	0.1140364	***	0.0032114	2.8161484
moses_intercal_feb_2019:ctd_cc1441004:temp	0.9926212	***	0.9562691 ***	-0.0020758	0.0767775	***	0.0061902	8.0624610
moses_intercal_feb_2019:diver_ufz_a5616:div_temp	0.9999524	***	1.0165609 ***	0.0005371	-0.0082756	***	0.0011634	14.0582718
moses_intercal_feb_2019:diver_ufz_a5644:div_temp	0.9999473	***	1.0169230 ***	0.0003625	0.0469419	***	0.0011941	2.5437916
moses_intercal_feb_2019:tid_ufz_10924450:tidb_temp	0.9999064	***	1.0038544 ***	-0.0006634	-0.0244756	***	0.0006742	1.4912897
moses_intercal_feb_2019:ctd_25564:Twater	0.9999425	***	1.0028887 ***	-0.0006723	-0.0244942	***	0.0005239	2.1399671
moses_intercal_feb_2019:ctd_25601:Twater	0.9999401	***	1.0032472 ***	-0.0000784	0.0991877	***	0.0005417	2.2118492
moses_intercal_feb_2019:tid_ufz_10669650:tidb_temp	0.9999089	***	1.0067264 ***	-0.0002065	-0.0323417	***	0.0007551	0.7612485
moses_intercal_feb_2019:exo1_geomar_0001:temperature_17f102065	0.9997965	***	0.9938749 ***	-0.0002065	-0.0323417	***	0.0012814	3.9620727
moses_intercal_feb_2019:ros_0001:temperature	0.0524740	ns	-69.8181698 ***	-0.0848775	-0.1951538	ns	0.3217956	164.8933121
moses_intercal_feb_2019:opus_geomar_71f9:temperature	0.9998785	***	0.9963838 *	0.0008304	-0.0207693	***	0.0021471	10.3380232
moses_intercal_feb_2019:ctd_37:16685	0.9999204	***	1.0010742 ***	-0.0000331	-0.0220024	***	0.0005576	1.4518944
moses_intercal_feb_2019:suna_geomar_345:temperature	0.9998968	***	0.9942930 ***	-0.0000331	-0.0220024	***	0.0020931	9.5131915
moses_intercal_feb_2019:fb_hzg_pocket0001:temperature_optode	0.9998817	***	0.9921091 ***	0.0011739	0.1568924	***	0.0008014	0.5107953

Unit price: 250 €

Unit price: 15 000 €

In ecology, data handling and verification procedures are by far not accurate and precise.

- Our data handling and verification procedures are only a first step towards a convincing quality control.
 - The state-of-the-art data handling procedures (flagging) are by far not sufficient for a high scientific level. What are probably good data?
 - How are missing data or data gaps filled?
 - We do not use state of the art capabilities of online sensor technology to countercheck data against other probes.
 - We do not use forecasting methods for online sensor control.
 - **We MUST provide Accuracy and Precision for each data point.**

We need more intelligent and automated data control and data verification procedures to achieve a higher data quality within a reasonable effort.

Cross-cutting problems



- Independent of the scientific discipline, the same problems emerge in almost any discussion.
 - Data processing with respect to data quality.
 - Stable data assessment in (near-realtime).
- Integration, aggregation & analysis of simple and complex data.

Integration & analysis of simple and complex data.

A real world example: Master thesis (##### University – Department Computational Science) on the effects of low water years on the water quality of the Southern North Sea.

The Task: To visualize existing MOSES data on the Elbe discharge on the water quality (Temperature, Salinity, Turbidity) in the Southern North Sea:

The target data sources :

- Cuxhaven Ferrybox, Helgoland Ferrybox... (TS data).
- Helgoland Reede Data (PANGAEA).
- German bight, FINO3, NSB II and NSB III (BSH Data Base).
- Cuxhaven Water Level (OPENDAP Data Base):
- River Discharge (Datenportal FGG Elbe):

A real world problem: The student needed about 2 month to successfully retrieve the data from the respective databases and to bring them in a form to use them for analysis:

- 11 R scripts with up to 1000 lines of code were necessary to convert the database data.

A non representative poll...



1. 70% of the biological scientists are not able to use R or Matlab...
2. 90% of the biologists are not prepared / trained to use NETcdf, JSON, XML etc. etc.
3. 90% of the scientist use Excell and standard calculation programs.
4. 90% of the ecological oriented biological scientists think that the term FAIR is related to fair trade products.

A real-world problem...

Access to data repositories like below is NOT feasible for most ecologists! -> (and will not be in the future):

http://sos.████████/sos.py?request=GetObservation&service=SOS&eventTime=2019-04-16T08:46:10Z/2019-04-16T10:08:10Z&offering=Prandtl&observedProperty=Sal_Teledyne_Citadel_TS-NH

https://dashboard.████████/data-xxl/rest/data?beginDate=2019-05-18T00:00:00&endDate=2019-05-19T23:59:59&format=application/json&aggregate=hour&aggregateFunctions=MEAN&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:pH_meinsberg_ega&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:relative_humidity_0001&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:salinity&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_aanderaa_optode&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_citadel&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_turner_scufa_0001&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:total_alcalinity_contros_fia_ta&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:turbidity&sensors=laboratory:moses_stern_1:fb_hzg_pocket0001:xco2_corr_contros_hydroc

The real world... 😊

Standard access of ecologists to the institutes ecological databases **IS** via “click and download”:

Available sensors

Filter sensors:

Sensor code	Sensor description		
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:pH_meinsberg_ega	pH	laboratory:moses_stern_1:fb_hzg_pocket0001:pH_meinsberg_ega	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:relative_humidity_0001	%	laboratory:moses_stern_1:fb_hzg_pocket0001:relative_humidity_0001	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:salinity	PSU	laboratory:moses_stern_1:fb_hzg_pocket0001:salinity	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_aanderaa_optode	°C	laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_aanderaa_optode	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_citadel	°C	laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_citadel	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_turner_scufa_0001	°C	laboratory:moses_stern_1:fb_hzg_pocket0001:temperature_turner_scufa_0001	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:total_alcalinity_contros_fia_ta	μmol/kg	laboratory:moses_stern_1:fb_hzg_pocket0001:total_alcalinity_contros_fia_ta	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:turbidity	FTU	laboratory:moses_stern_1:fb_hzg_pocket0001:turbidity	
<input checked="" type="checkbox"/> laboratory:moses_stern_1:fb_hzg_pocket0001:xco2_corr_contros_hydroc	μatm	laboratory:moses_stern_1:fb_hzg_pocket0001:xco2_corr_contros_hydroc	

9 / 1284 sensor(s) are registered for this data service.

Request data

Begin
 End
 Format
 Aggregate
 minimum
 0.25-percentile
 mean
 median
 0.75-percentile
 maximum
 standard deviation
 count

We often discuss that we need a good public outreach and that we must provide our data, if possible open access, to the relevant “Stakeholders”.

We fully agree with this 😊

but

With respect to data, ecological scientists, which often do not have much computational competences in data-mining and large-scale data exploration, are most important stakeholders using the data for real science.