FAIR research data for agrosystems
Driver for soil health and sustainable crop production

Carsten Hoffmann

Leibniz Centre for Agricultural Landscape Research (ZALF)

Muencheberg, Germany

Leibniz-Zentrum für Agrarlandschaftsforschung (ZALF) e.V.

Data Scientist BonaRes project

BONARES Centre for Soil Research

Sponsored by the Federal Ministry of Education and Research

Project Manager FAIRagro project

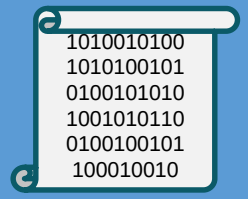
FAIRagro

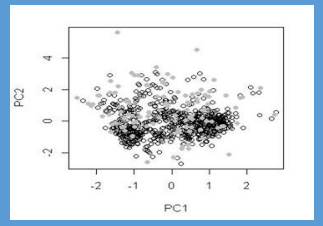
Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

Imagine…

USE CASE 1
(researcher)

you have finished a long measure campain in the field and in the lab;





spent many hours in cleaning, preparing and analysing your data;

spent even more time in writing a paper,

spent even more energy until the paper was finally published in a high-ranking journal;

can’t wait to hear your colleagues’ responses and be cited…….but….

something is missing;

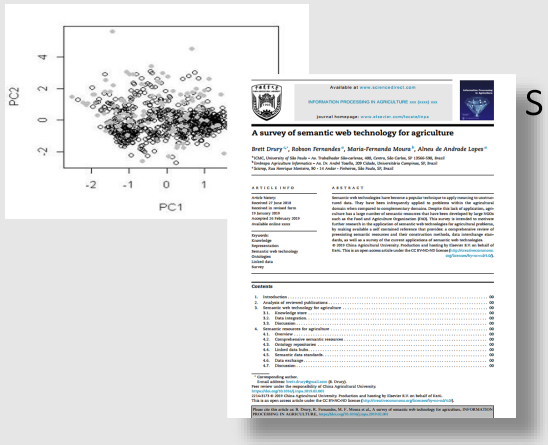
Whats about your valuable datasets? Are they findable, accesible and reusable for your colleagues?

Wouldn‘t it be fair to spread your findings and increase your and your research visibility?

USE CASE 2
(modeller for a food company)

….and imagine

your task is to model soil fertility and a crop scenario in a certain region;



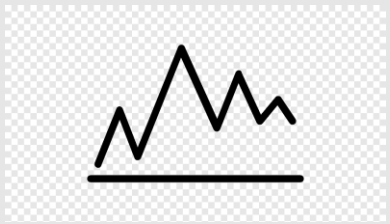
spent many hours in reading papers, extract tables and figure values to feed you models;

but the data are too patchy and the uncertainty is too high for good predictions;

you try to get in contact with (paper) authors to get access to measured data;

some authors vanished, other respond an send you data in different formats, file types described by none /sparse metadata;

finally you give up and work with low quality model outcomes.



Wouldn‘t it be fair to get permanent access to data and base knowledge from public research?

The FAIR data principles can help…

Published soil- and agricultural research data, should ….

- ✓ be visible and findable for others
- ✓ be accessible
- ✓ be legally sound (authors rights respected)
- ✓ aggregates data usefully
- ✓ be quality assured, and
- ✓ be internationally harmonized, and
- ✓ reusable in a user-friendly way

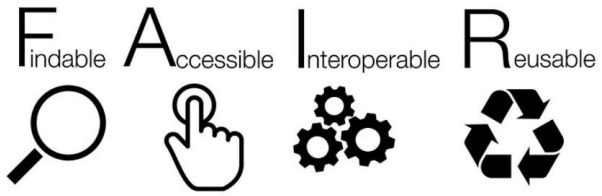
Findable

Accessible

Interoperable

Reusable

→ FAIR Data Principles



Standards

Open community standards
(freely available and widely accepted)

- ✓ Metadata elements (e.g. schema.org, DataCite)
- ✓ PIDs (e.g. ORCID, RoR)
- ✓ Thesauri (e.g. AGROVOC)
- ✓ Geodata (OGC)
- ✓ Crop & variety codes (e.g. ICC, FAO)
- ✓ Soil classification (WRB)
- ✓ Methods (e.g. ISO)
- ✓ Licencing (Creative Commons)

→ FAIR Data Principles

Findable Accessible Interoperable Reusable

ISOBUS

How to handle field data?

Most important: Metadata!

	A	B	C	D	E	F	G	H	I	J	K
1	Sites	Year	Dates	Plot_Nr	Landuse	Plot	CO2	CH4	WFPS	Soil_Temp	NH4
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	DR1	16,15248907	2,014537418	66,3090727	2,8	0,353873015
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	DR2	21,57602373	-17,58430694	70,1414122	3,2	0,27466504
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	DR3	14,23285353	-6,55049326	60,0252173	3,9	0,323489214
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	DR4	7,668587595	1,369836777	71,1751839	3,8	0,151494886
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	WT1	17,38527193	-9,029799495	67,4452373	5,2	3,629601916
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	WT2	26,09982689	-2,861573591	65,086966	8,5	7,95840662
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	WT3	23,7712574	-6,790432495	60,4034525	5,3	4,797791461
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	WT4	64,91581054	10,12928832	76,5546381	7,2	6,295071121
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	WA1	24,20015152	-3,926820399	54,907974	5,6	10,99501079
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	WA2	24,14647683	-5,825871517	49,639882	8,5	4,156102935
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	WA3	14,63899374	3,577959891	60,1435066	5,8	2,551189552
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	WA4	47,62797412	-11,38352577	41,1809805	9,8	14,50451548

Metadata

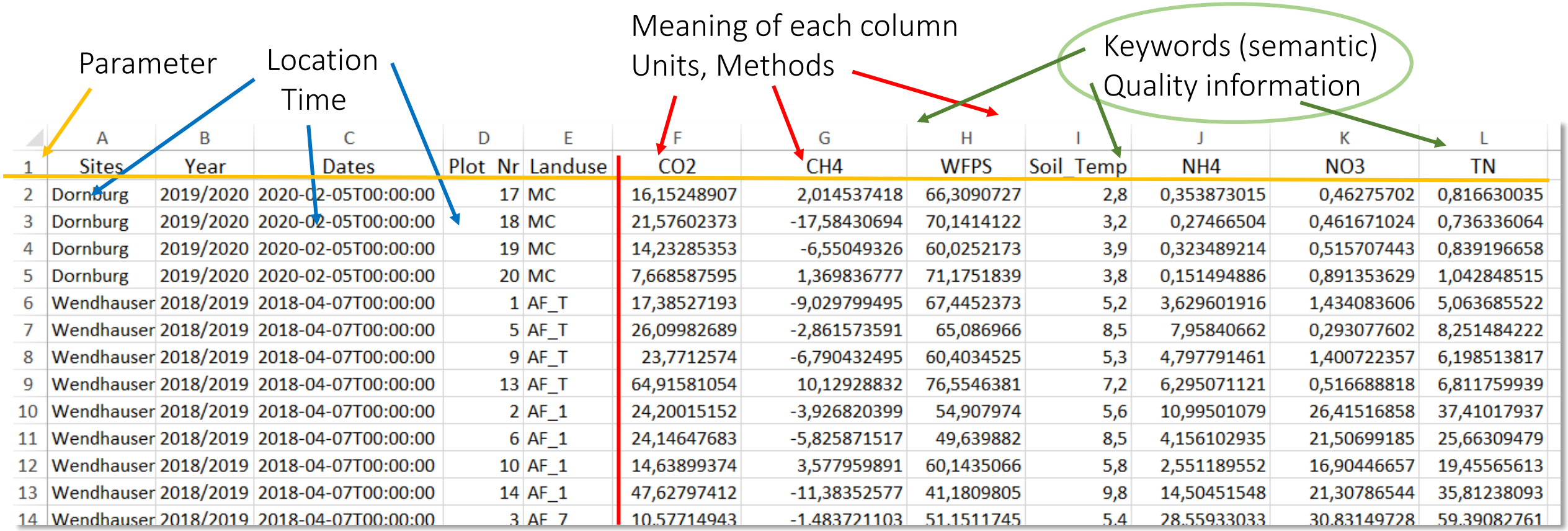
= “data about data”, your data label

- **essential ”partners”** for field and research data
- Link to keywords, ontologies
- **Standardised** and structured information
- Subset of **documentation**: describes, explains, locates, makes it easier to retrieve, use, manage an information resource
- **Human-** and **machine-readable**

WHO generated the data?
HOW was it generated and processed?
WHAT is the content of the data?
WHY was the data generated?
WHEN was the data generated?
WHERE was the data generated?

How to handle field data?

Metadata: what does the data mean? (the whole table but also each column):



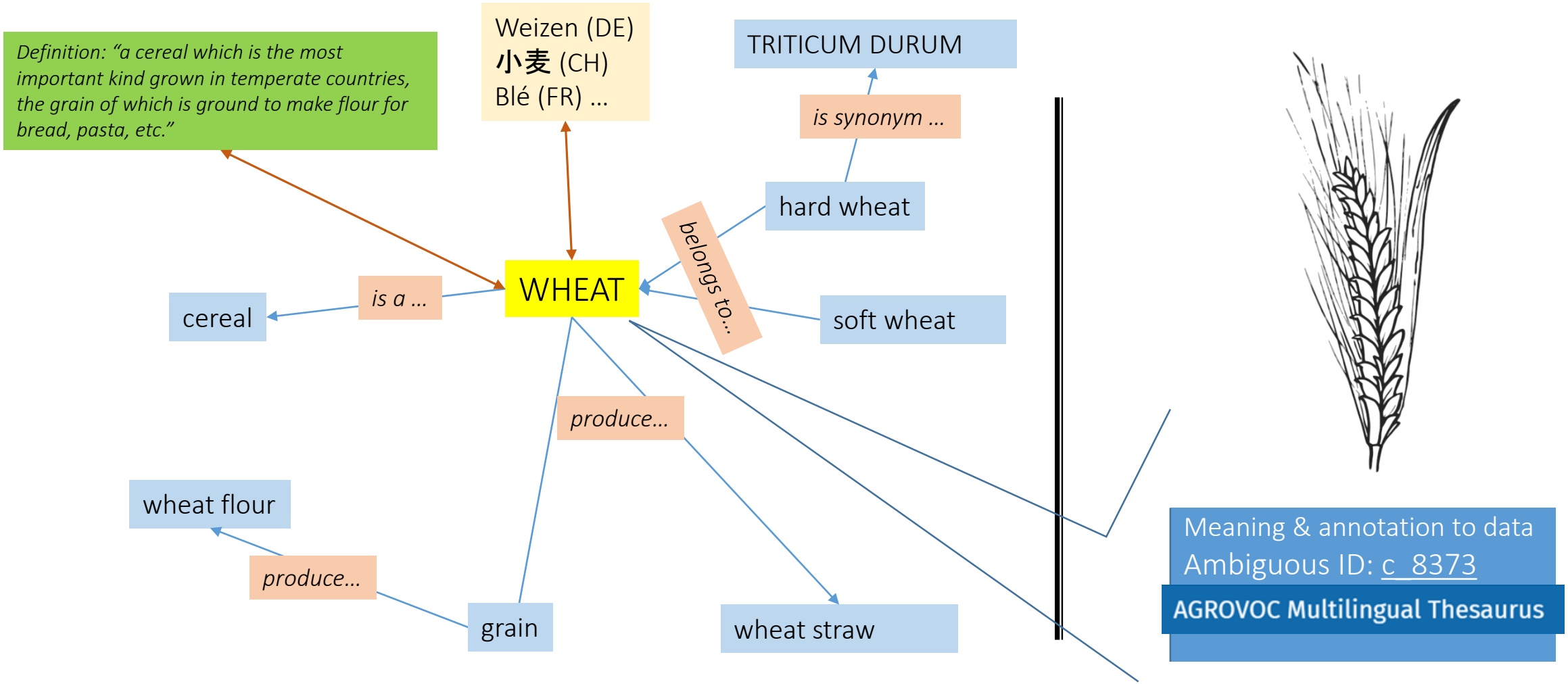
	A	B	C	D	E	F	G	H	I	J	K	L
1	Sites	Year	Dates	Plot_Nr	Landuse	CO2	CH4	WFPS	Soil_Temp	NH4	NO3	TN
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	16,15248907	2,014537418	66,3090727	2,8	0,353873015	0,46275702	0,816630035
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	21,57602373	-17,58430694	70,1414122	3,2	0,27466504	0,461671024	0,736336064
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	14,23285353	-6,55049326	60,0252173	3,9	0,323489214	0,515707443	0,839196658
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	7,668587595	1,369836777	71,1751839	3,8	0,151494886	0,891353629	1,042848515
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	17,38527193	-9,029799495	67,4452373	5,2	3,629601916	1,434083606	5,063685522
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	26,09982689	-2,861573591	65,086966	8,5	7,95840662	0,293077602	8,251484222
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	23,7712574	-6,790432495	60,4034525	5,3	4,797791461	1,400722357	6,198513817
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	64,91581054	10,12928832	76,5546381	7,2	6,295071121	0,516688818	6,811759939
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	24,20015152	-3,926820399	54,907974	5,6	10,99501079	26,41516858	37,41017937
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	24,14647683	-5,825871517	49,639882	8,5	4,156102935	21,50699185	25,66309479
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	14,63899374	3,577959891	60,1435066	5,8	2,551189552	16,90446657	19,45565613
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	47,62797412	-11,38352577	41,1809805	9,8	14,50451548	21,30786544	35,81238093
14	Wendhauser	2018/2019	2018-04-07T00:00:00	3	AF_7	10,57714943	-1,483721103	51,1511745	5,4	28,55933033	30,83149728	59,39082761

+ general: research question, provenance, authors contact……

https://agrovoc.fao.org/browse/agrovoc/en/

Semantics:
Teach a machine (search engine): What is „wheat“?

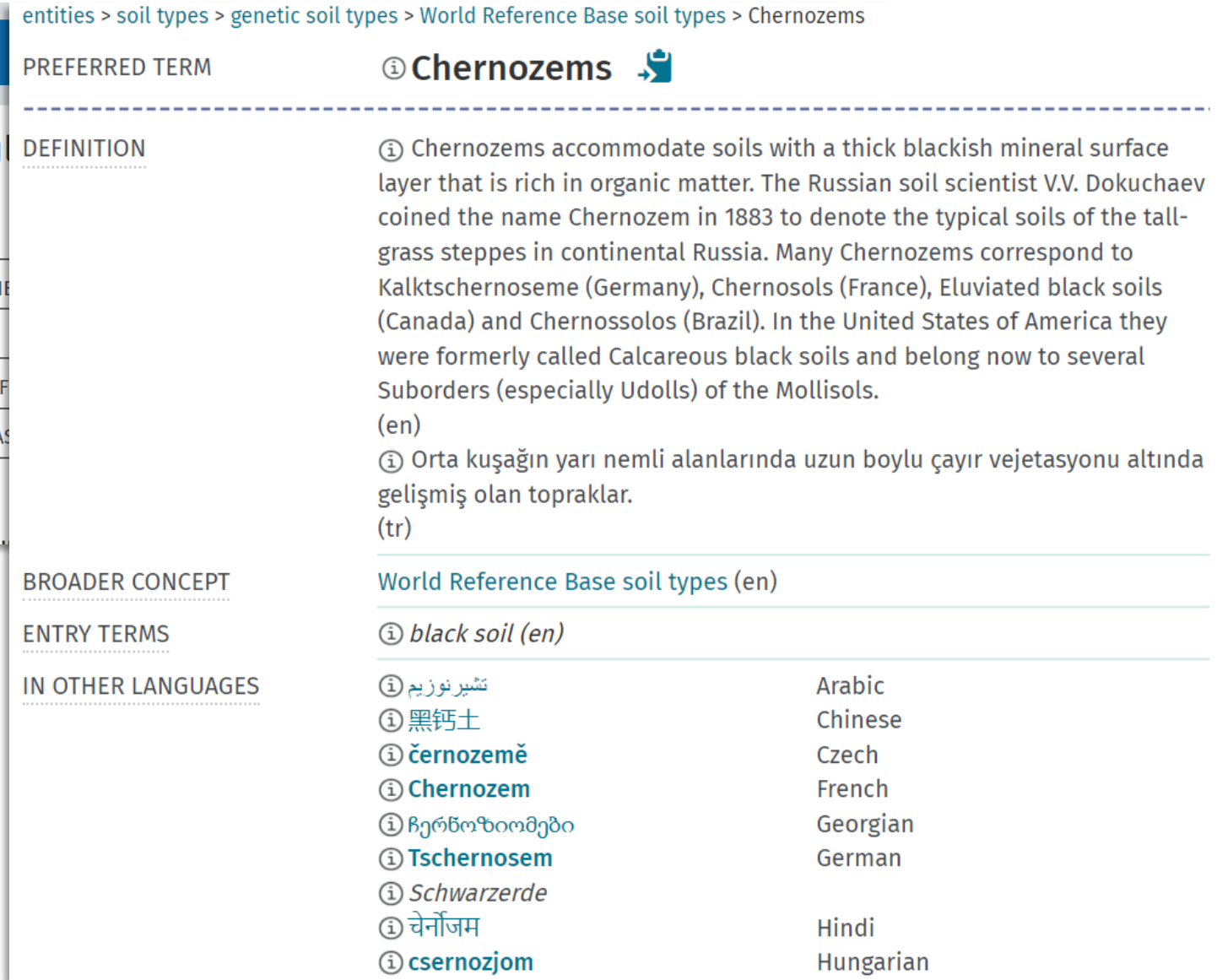
AGROVOC Multilingual Thesaurus



https://agrovoc.fao.org/browse/agrovoc/en/

provide „controlled“ keywords

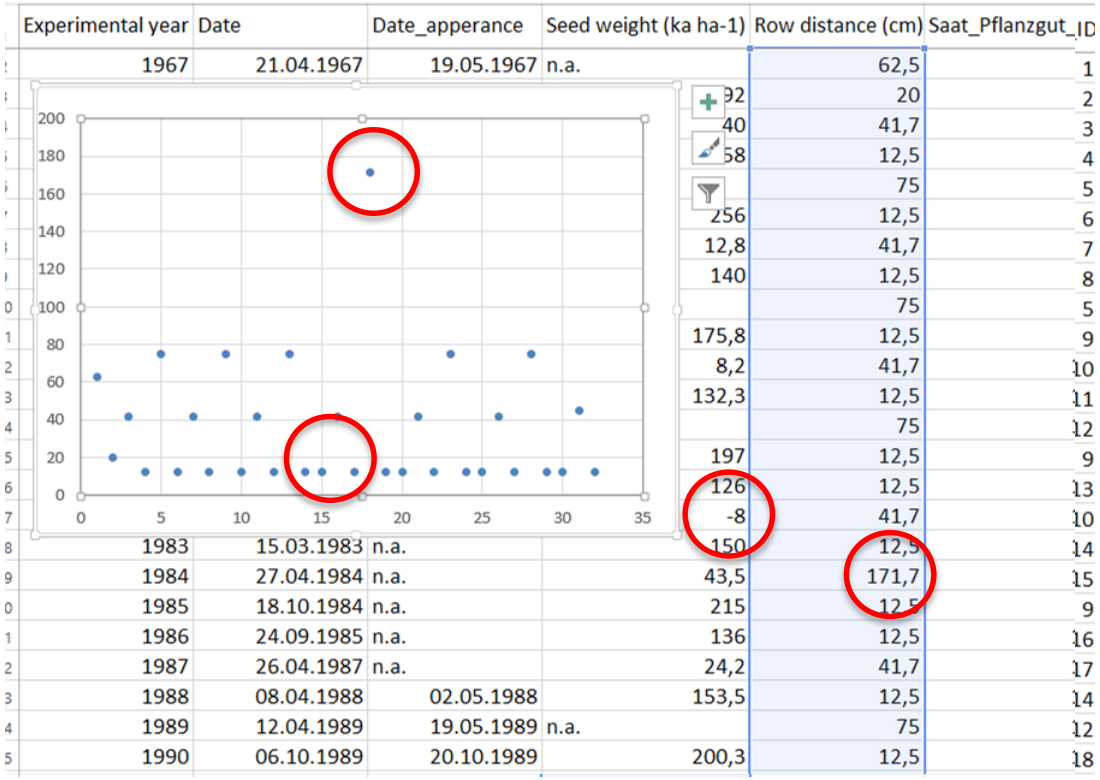
AGROVOC Multilingual Thesaurus



AGROVOC
operated by FAO, open access
~42.000 concepts (terms)
>100.000 alternative concepts
~45 languages
machine readable

Data Quality

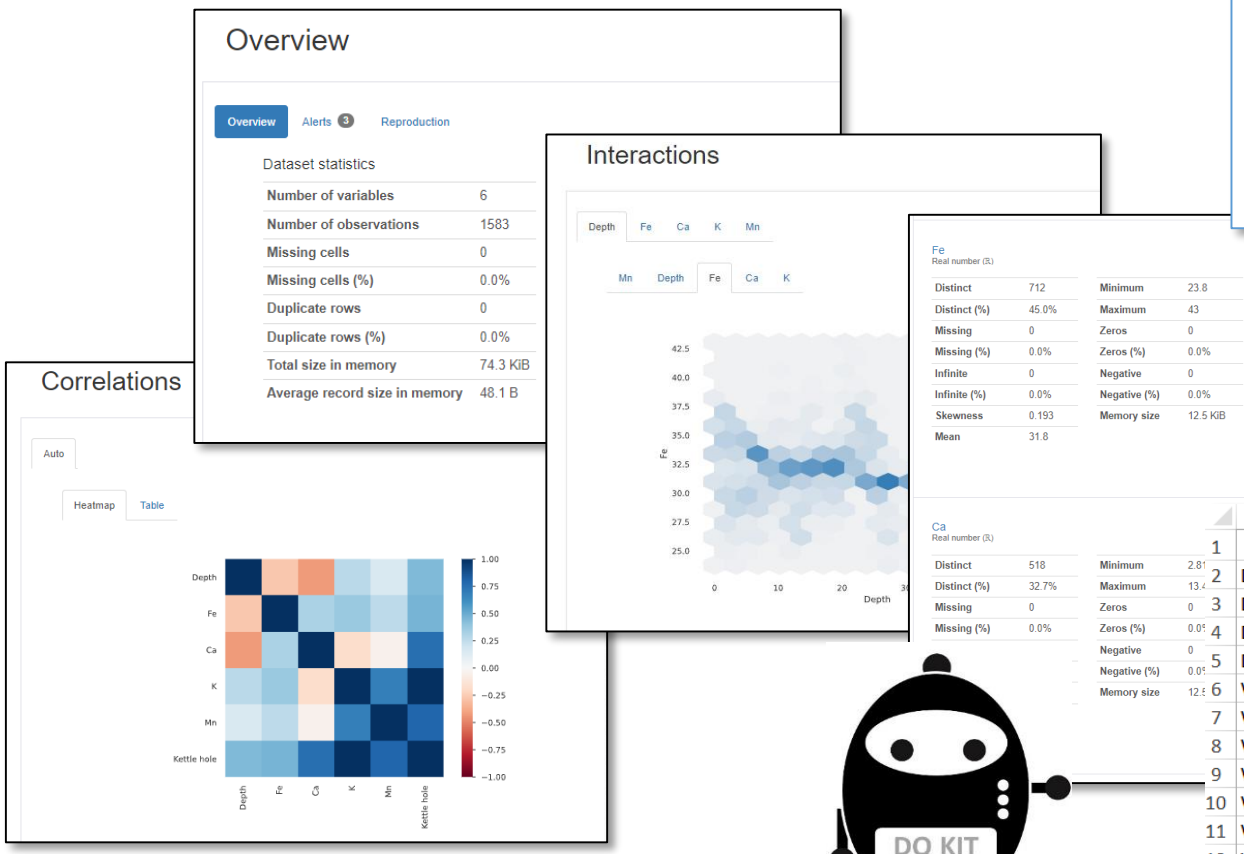
- trustworthiness of a dataset,
- **formal** requirements (e.g. sql conformity),
- **content-related** requirements (e.g. plausibility, statistically sound, complete, …)
- test before data upload by data provider /plausibility DB

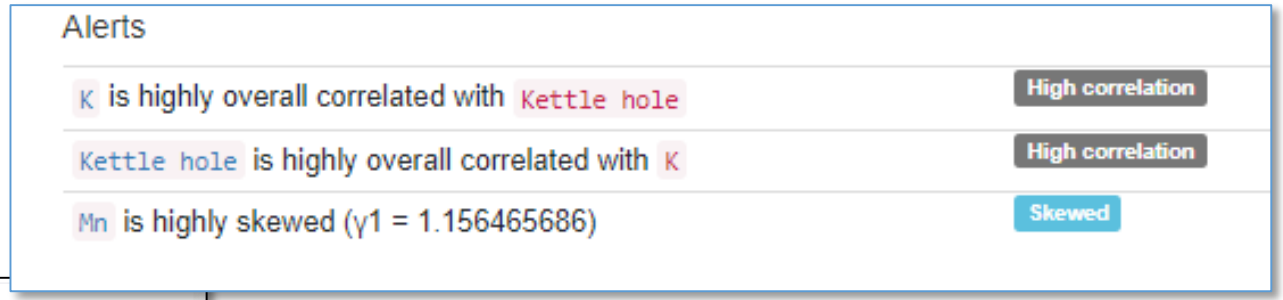


Experimental year	Date	Date_apperance	Seed weight (ka ha-1)	Row distance (cm)	Saat_Pflanzgut_ID
1967	21.04.1967	19.05.1967	n.a.	62,5	1
			92	20	2
			40	41,7	3
			58	12,5	4
				75	5
			256	12,5	6
			12,8	41,7	7
			140	12,5	8
				75	5
			175,8	12,5	9
			8,2	41,7	10
			132,3	12,5	11
				75	12
			197	12,5	9
			126	12,5	13
			-8	41,7	10
1983	15.03.1983	n.a.	150	12,5	14
1984	27.04.1984	n.a.	43,5	171,7	15
1985	18.10.1984	n.a.	215	12,5	9
1986	24.09.1985	n.a.	136	12,5	16
1987	26.04.1987	n.a.	24,2	41,7	17
1988	08.04.1988	02.05.1988	153,5	12,5	14
1989	12.04.1989	19.05.1989	n.a.	75	12
1990	06.10.1989	20.10.1989	200,3	12,5	18

BONARES
Centre for Soil Research

Data quality (DQ KIT)





- pdf download DQ report (supplement. material)
- results to be stored in metadata

	A	B	C	D	E	F	G	H	I	J	K
1	Sites	Year	Dates	Plot_Nr	Landuse	Plot	CO2	CH4	WFPS	Soil_Temp	NH4
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	DR1	16,15248907	2,014537418	66,3090727	2,8	0,3538730
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	DR2	21,57602373	-17,58430694	70,1414122	3,2	0,274665
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	DR3	14,23285353	-6,55049326	60,0252173	3,9	0,3234892
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	DR4	7,668587595	1,369836777	71,1751839	3,8	0,1514948
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	WT1	17,38527193	-9,029799495	67,4452373	5,2	3,6296019
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	WT2	26,09982689	-2,861573591	65,086966	8,5	7,958406
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	WT3	23,7712574	-6,790432495	60,4034525	5,3	4,7977914
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	WT4	64,91581054	10,12928832	76,5546381	7,2	6,2950711
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	WA1	24,20015152	-3,926820399	54,907974	5,6	10,995010
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	WA2	24,14647683	-5,825871517	49,639882	8,5	4,1561029
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	WA3	14,63899374	3,577959891	60,1435066	5,8	2,5511895
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	WA4	47,62797412	-11,38352577	41,1809805	9,8	14,504515

dqkit.bonares.de

BONARES
Centre for Soil Research

How to handle field data?

	A	B	C	D	E	F	G	H	I	J	K
1	Sites	Year	Dates	Plot_Nr	Landuse	Plot	CO2	CH4	WFPS	Soil_Temp	NH4
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	DR1	16,15248907	2,014537418	66,3090727	2,8	0,353873015
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	DR2	21,57602373	-17,58430694	70,1414122	3,2	0,27466504
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	DR3	14,23285353	-6,55049326	60,0252173	3,9	0,323489214
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	DR4	7,668587595	1,369836777	71,1751839	3,8	0,151494886
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	WT1	17,38527193	-9,029799495	67,4452373	5,2	3,629601916
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	WT2	26,09982689	-2,861573591	65,086966	8,5	7,95840662
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	WT3	23,7712574	-6,790432495	60,4034525	5,3	4,797791461
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	WT4	64,91581054	10,12928832	76,5546381	7,2	6,295071121
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	WA1	24,20015152	-3,926820399	54,907974	5,6	10,99501079
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	WA2	24,14647683	-5,825871517	49,639882	8,5	4,156102935
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	WA3	14,63899374	3,577959891	60,1435066	5,8	2,551189552
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	WA4	47,62797412	-11,38352577	41,1809805	9,8	14,50451548

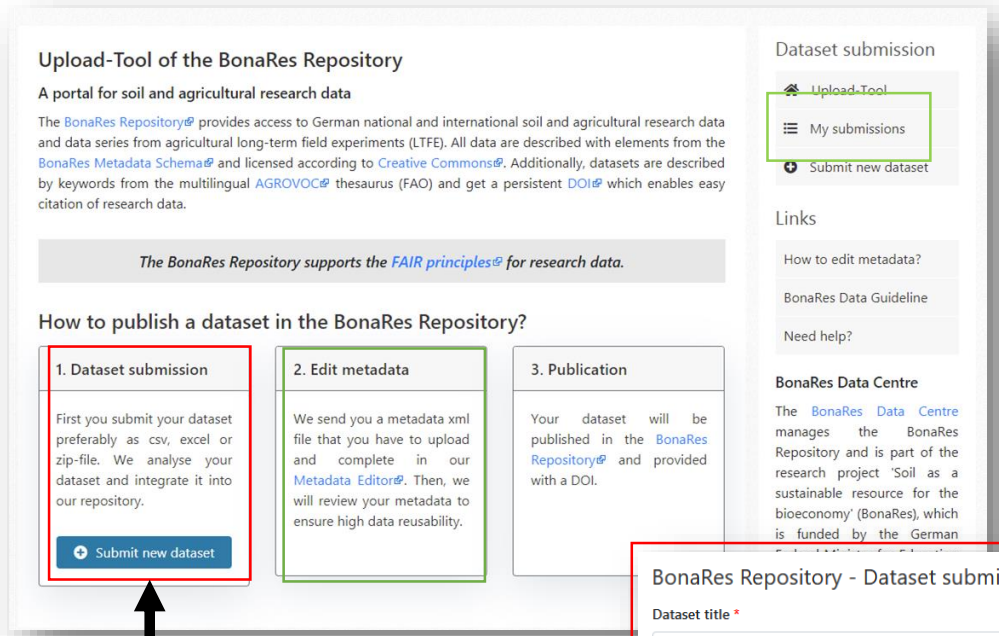
- ✓ described by metadata
- ✓ annotated controlled keywords
- ✓ **data quality approved**

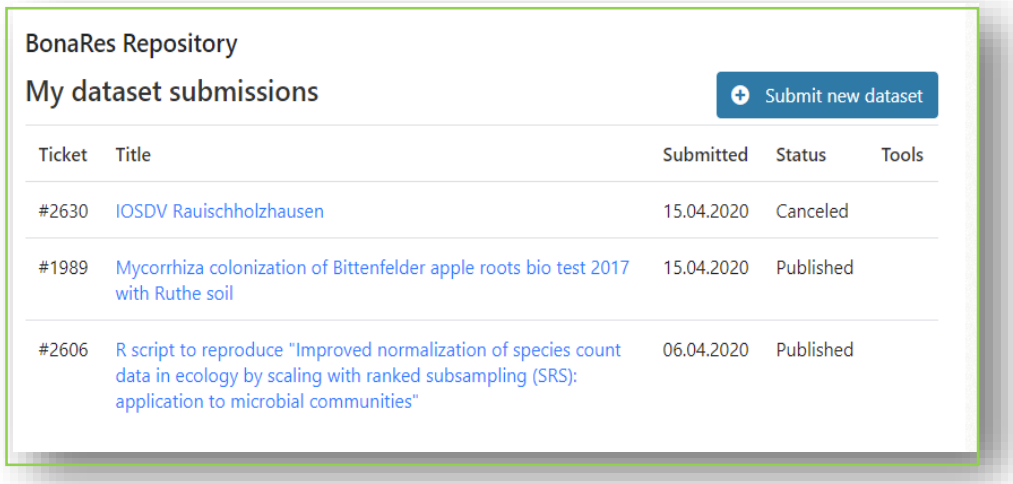
Data submission (BonaRes Repository)

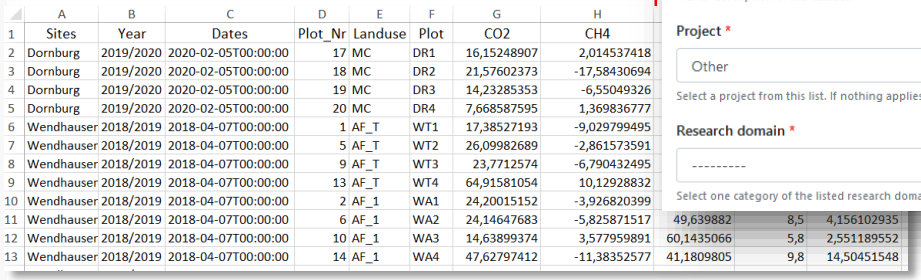
Upload tool

- Submission of QA and structured data
- Add metadata
- Easy editorial (ticket) system → status of submitted data
- Feedback (Data Stewards)
- Published (DOI!, FAIR!)

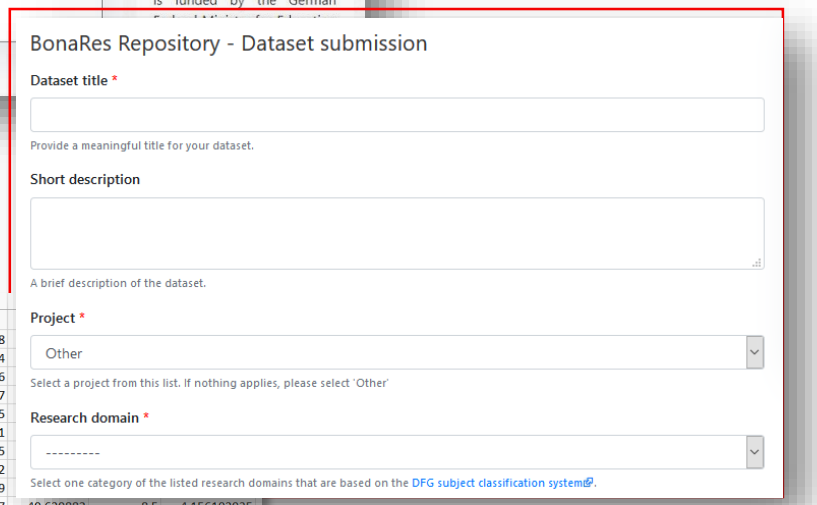
upload.bonares.de







	A	B	C	D	E	F	G	H			
1	Sites	Year	Dates	Plot_Nr	Landuse	Plot	CO2	CH4			
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	DR1	16,15248907	2,014537418			
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	DR2	21,57602373	-17,58430694			
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	DR3	14,23285353	-6,55049326			
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	DR4	7,668587595	1,369836777			
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	WT1	17,38527193	-9,029799495			
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	WT2	26,09982689	-2,861573591			
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	WT3	23,7712574	-6,790432495			
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	WT4	64,91581054	10,12928832			
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	WA1	24,20015152	-3,926820399			
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	WA2	24,14647683	-5,825871517	49,639882	8,5	4,156102935
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	WA3	14,63899374	3,577959891	60,1435066	5,8	2,551189552
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	WA4	47,62797412	-11,38352577	41,1809805	9,8	14,50451548



BONARES
Centre for Soil Research

Data publication (BonaRes Repository)

The powerful DOI to
connect data and paper

A survey of semantic web technology for agriculture

Brett Drury, Robson Fernandes, Maria-Fernanda Moura, Alneu de Andrade Lopes

doi.org/10.20387/BonaRes-PDY6-HHGS

Data

Paper

doi.org/10.5194/soil-4-23-2018

	A	B	C	D	E	F	G	H	I	J	K
1	Sites	Year	Dates	Plot_Nr	Landuse	Plot	CO2	CH4	WFPS	Soil_Temp	NH4
2	Dornburg	2019/2020	2020-02-05T00:00:00	17	MC	DR1	16,15248907	2,014537418	66,3090727	2,8	0,353873015
3	Dornburg	2019/2020	2020-02-05T00:00:00	18	MC	DR2	21,57602373	-17,58430694	70,1414122	3,2	0,27466504
4	Dornburg	2019/2020	2020-02-05T00:00:00	19	MC	DR3	14,23285353	-6,55049326	60,0252173	3,9	0,323489214
5	Dornburg	2019/2020	2020-02-05T00:00:00	20	MC	DR4	7,668587595	1,369836777	71,1751839	3,8	0,151494886
6	Wendhauser	2018/2019	2018-04-07T00:00:00	1	AF_T	WT1	17,38527193	-9,029799495	67,4452373	5,2	3,629601916
7	Wendhauser	2018/2019	2018-04-07T00:00:00	5	AF_T	WT2	26,09982689	-2,861573591	65,086966	8,5	7,95840662
8	Wendhauser	2018/2019	2018-04-07T00:00:00	9	AF_T	WT3	23,7712574	-6,790432495	60,4034525	5,3	4,797791461
9	Wendhauser	2018/2019	2018-04-07T00:00:00	13	AF_T	WT4	64,91581054	10,12928832	76,5546381	7,2	6,295071121
10	Wendhauser	2018/2019	2018-04-07T00:00:00	2	AF_1	WA1	24,20015152	-3,926820399	54,907974	5,6	10,99501079
11	Wendhauser	2018/2019	2018-04-07T00:00:00	6	AF_1	WA2	24,14647683	-5,825871517	49,639882	8,5	4,156102935
12	Wendhauser	2018/2019	2018-04-07T00:00:00	10	AF_1	WA3	14,63899374	3,577959891	60,1435066	5,8	2,551189552
13	Wendhauser	2018/2019	2018-04-07T00:00:00	14	AF_1	WA4	47,62797412	-11,38352577	41,1809805	9,8	14,50451548

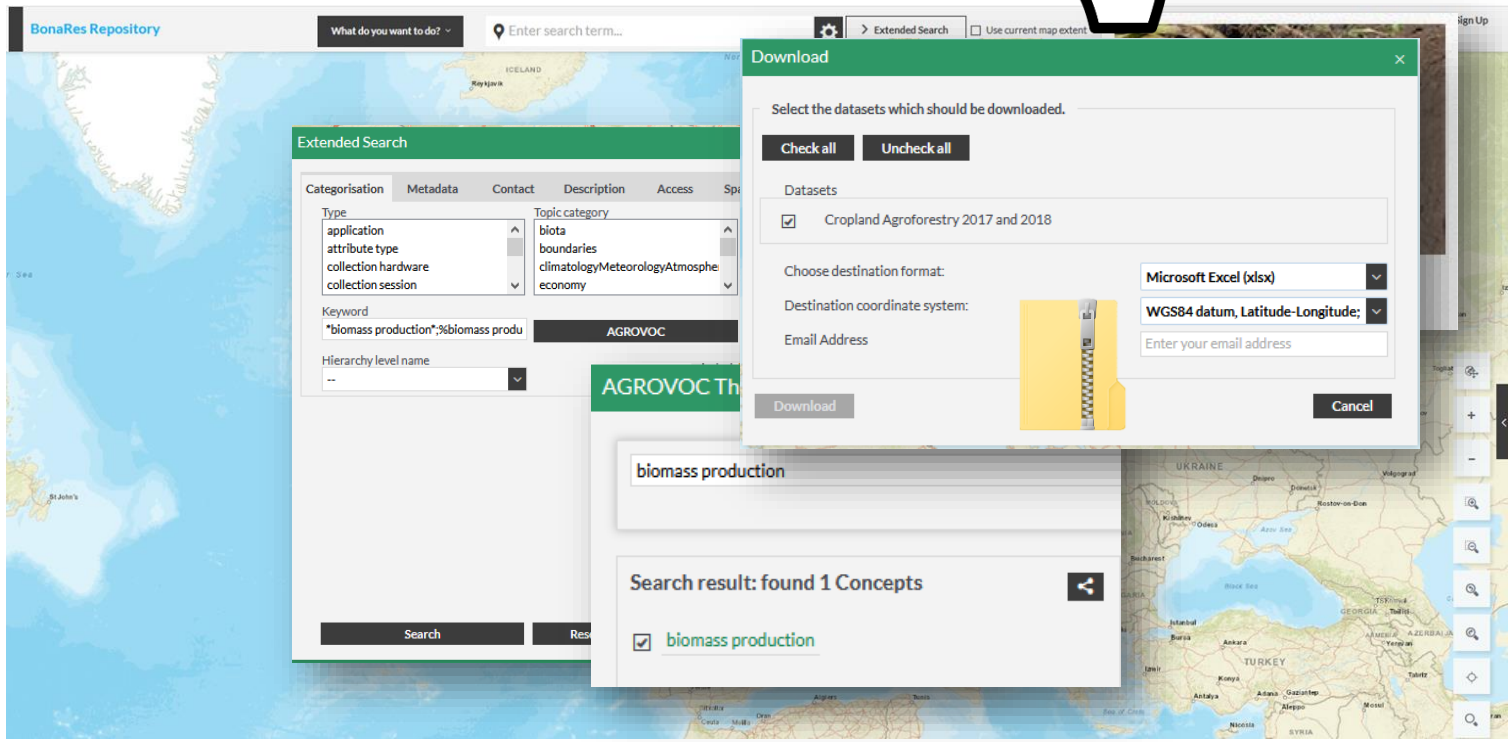
- ✓ described by metadata
- ✓ annotated controlled keywords
- ✓ data quality approved
- ✓ **published/accessible (DOI)**

BONARES
Centre for Soil Research

Search and find FAIR and quality assured soil/agri data

Field data: maps.bonares.de/mapapps/?lang=en
Long-term experiments: lte.bonares.de
Soil profiles: tools.bonares.de/bp_db/

- Different search options (map, AGROVOC keywords; DataCite,)
- Different download formats and coordinate systems
- No registration needed
- Rich metadata (CC-0; as XML or pdf)
- Re-use research data – open (CC-BY)



BONARES
Centre for Soil Research

NFDI – connect database „silos“

NFDI (=National Research Data Infrastructure)

- 28 consortia (from Agriculture to Zoology)
- cooperation not competion
- ethical aspects, industry involvement…
- 10 years financed, up to 900 M€
- national response to commercial “data players”, e.g. Google
- international harmonization

Nationale Forschungsdaten Infrastruktur

FAIRagro – connect agricultural database „silos“

Soil-Agrosystem data networks

FAO, AgMIP, MACSUR, GAIA-X, BONARES Centre for Soil Research, GL TEN Global Long-Term Agricultural Experiment Network, elixir, GBIF Global Biodiversity Information Facility, de.NBI, NCBI GenBank, GODAN Global Open Data for Agriculture & Nutrition, CGIAR, BrAPI, ISMC International Soil Modelling Consortium, AgroLD, DAKIS

Research data management (RDM) networks

W3C, OGC Making location count., GO FAIR, EUROPEAN OPEN SCIENCE CLOUD, ISO, RDA RESEARCH DATA ALLIANCE, OSGeo, FAIRsharing, DataCite FIND, ACCESS, AND REUSE DATA

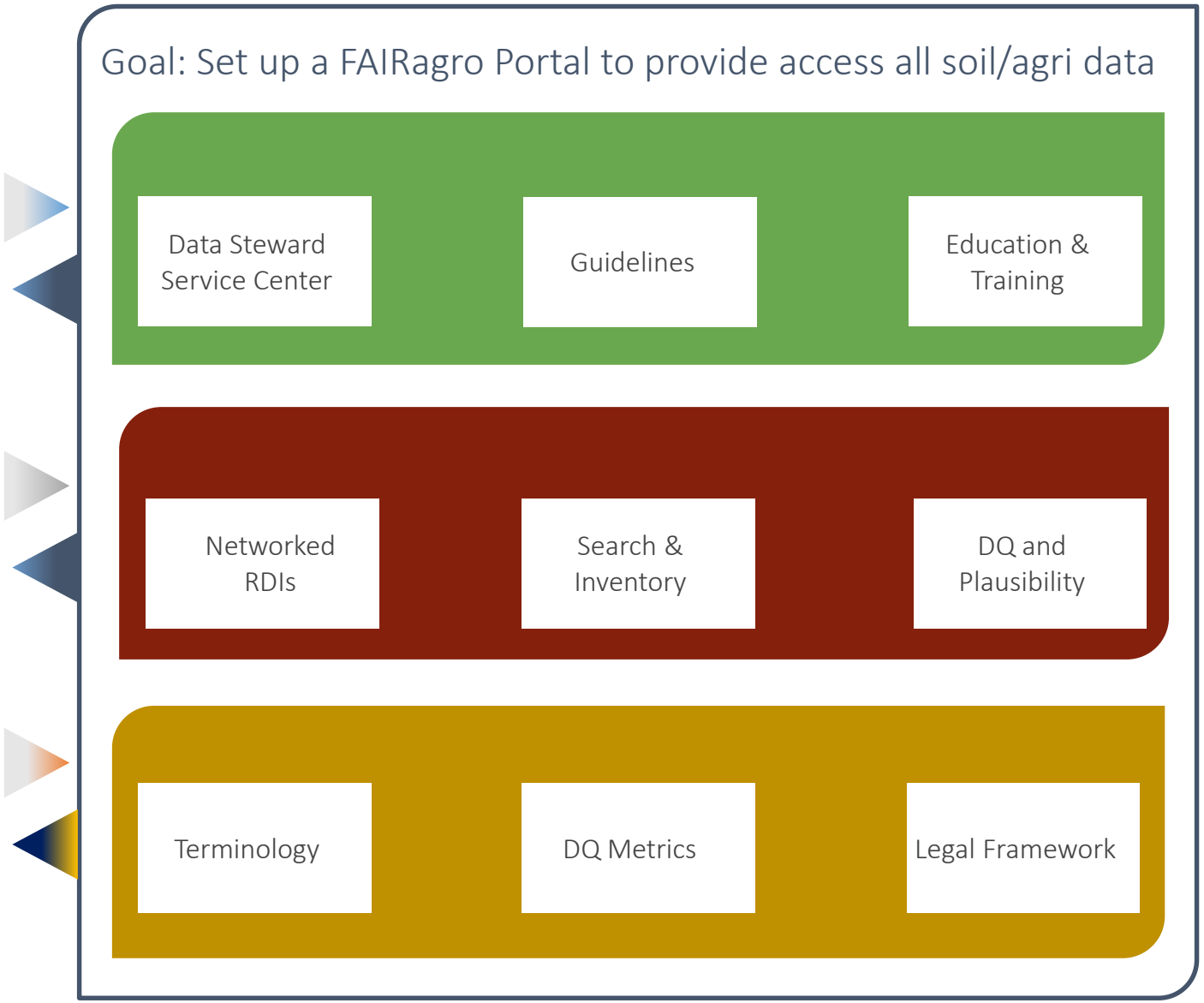
2023 - 2028

FAIRagro

FAIRagro – organisation

29 national partners (universities, research institutes, associations…)

Agrosystem Community
(what are your demands?)



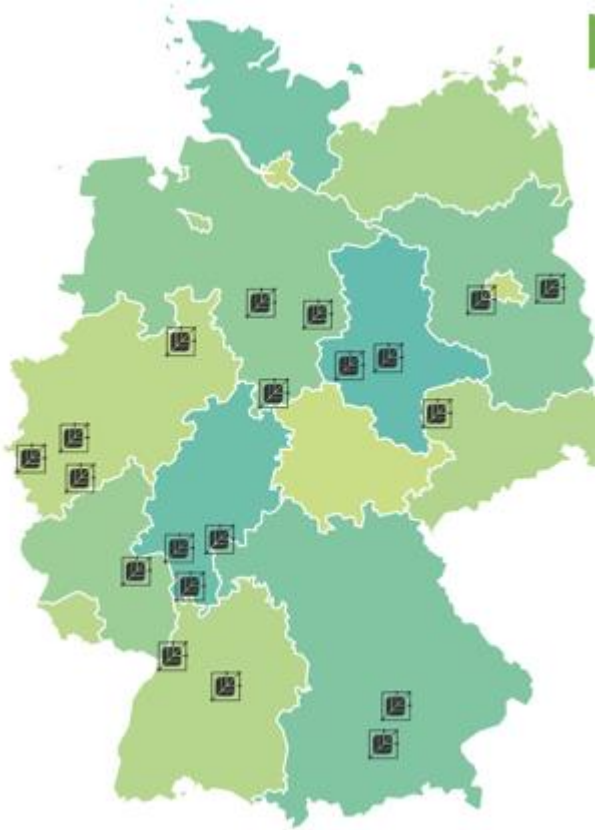
Support & Training

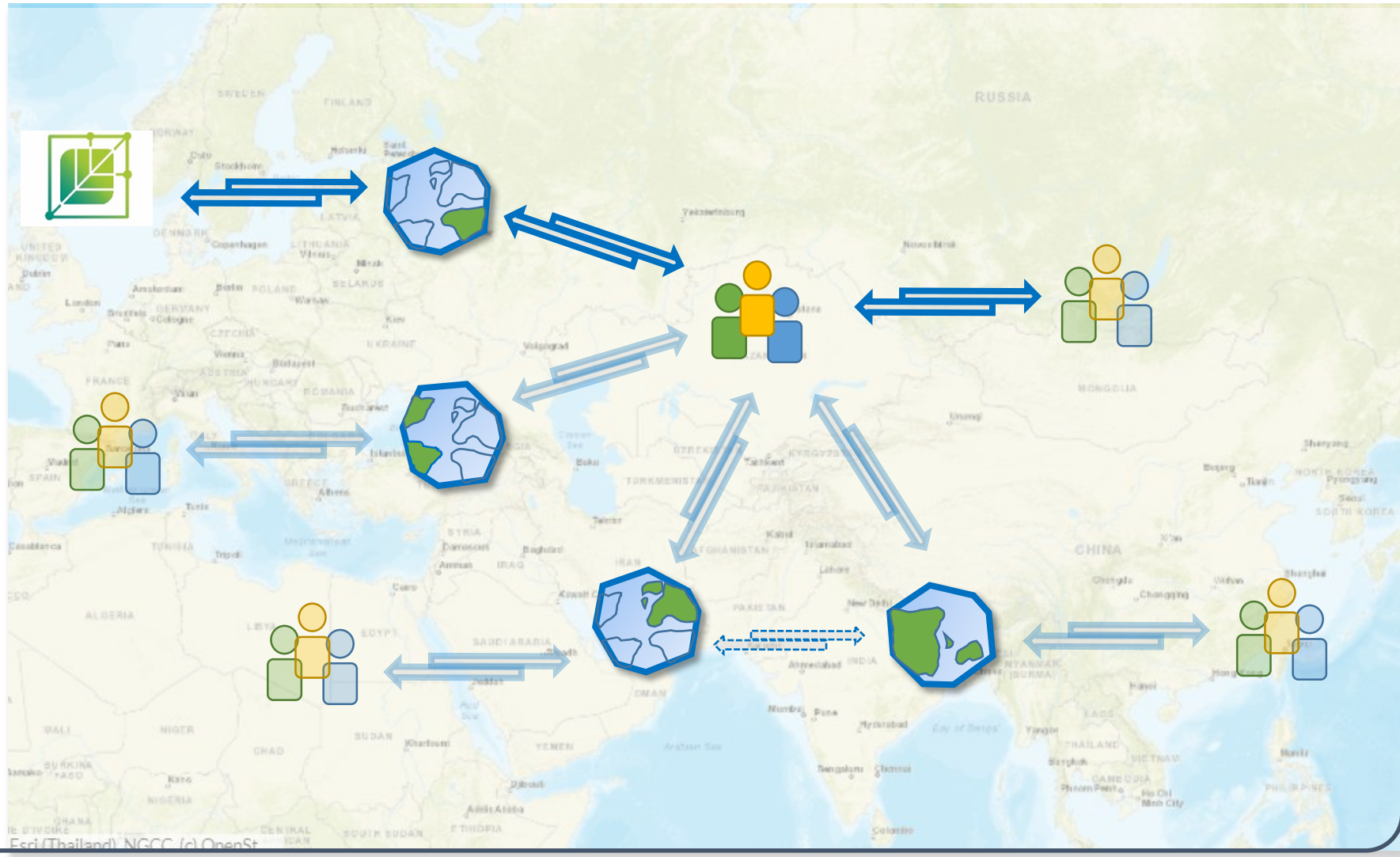
IT Services and Standards

FAIRagro

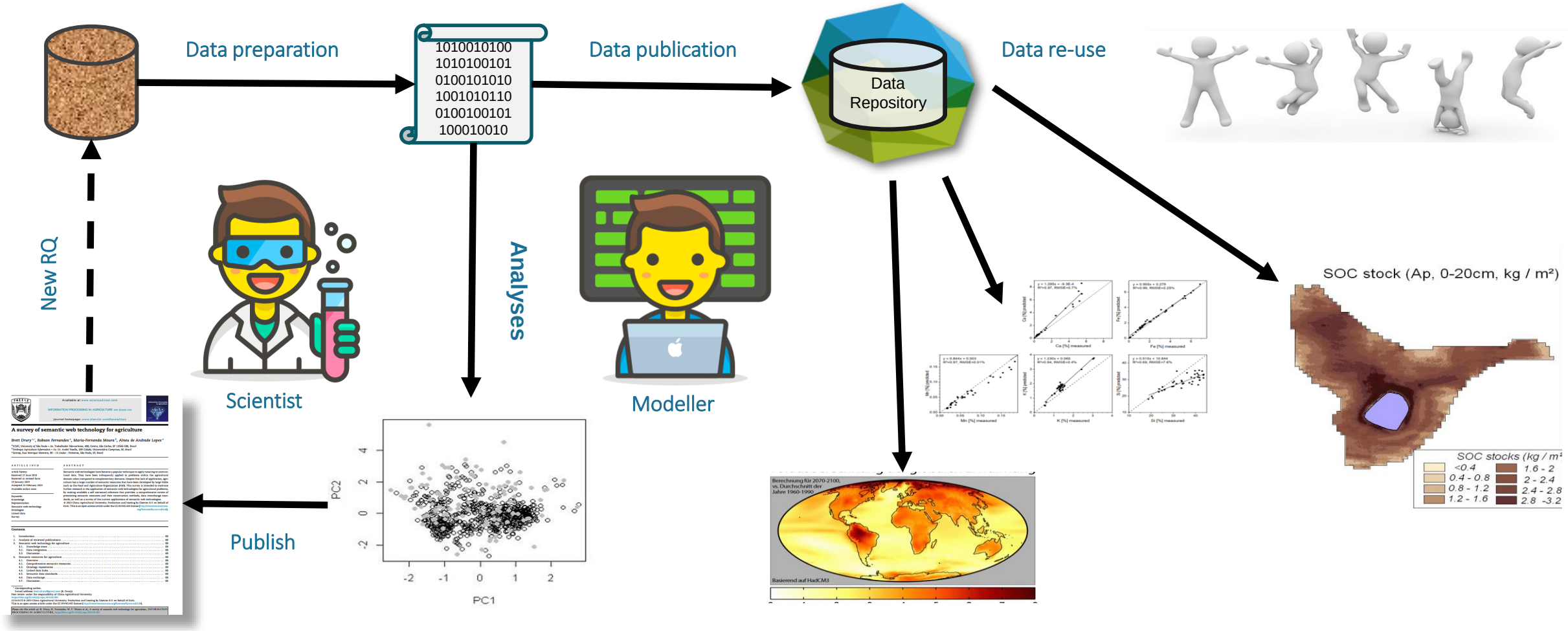
FAIRagro – One Portal for all

Goal: Set up a FAIRagro Portal to provide access all soil/agri data





Our use case from the beginning…



Thank you
Danke
Xie xie

for your attention!

(Black soil in South Germany)

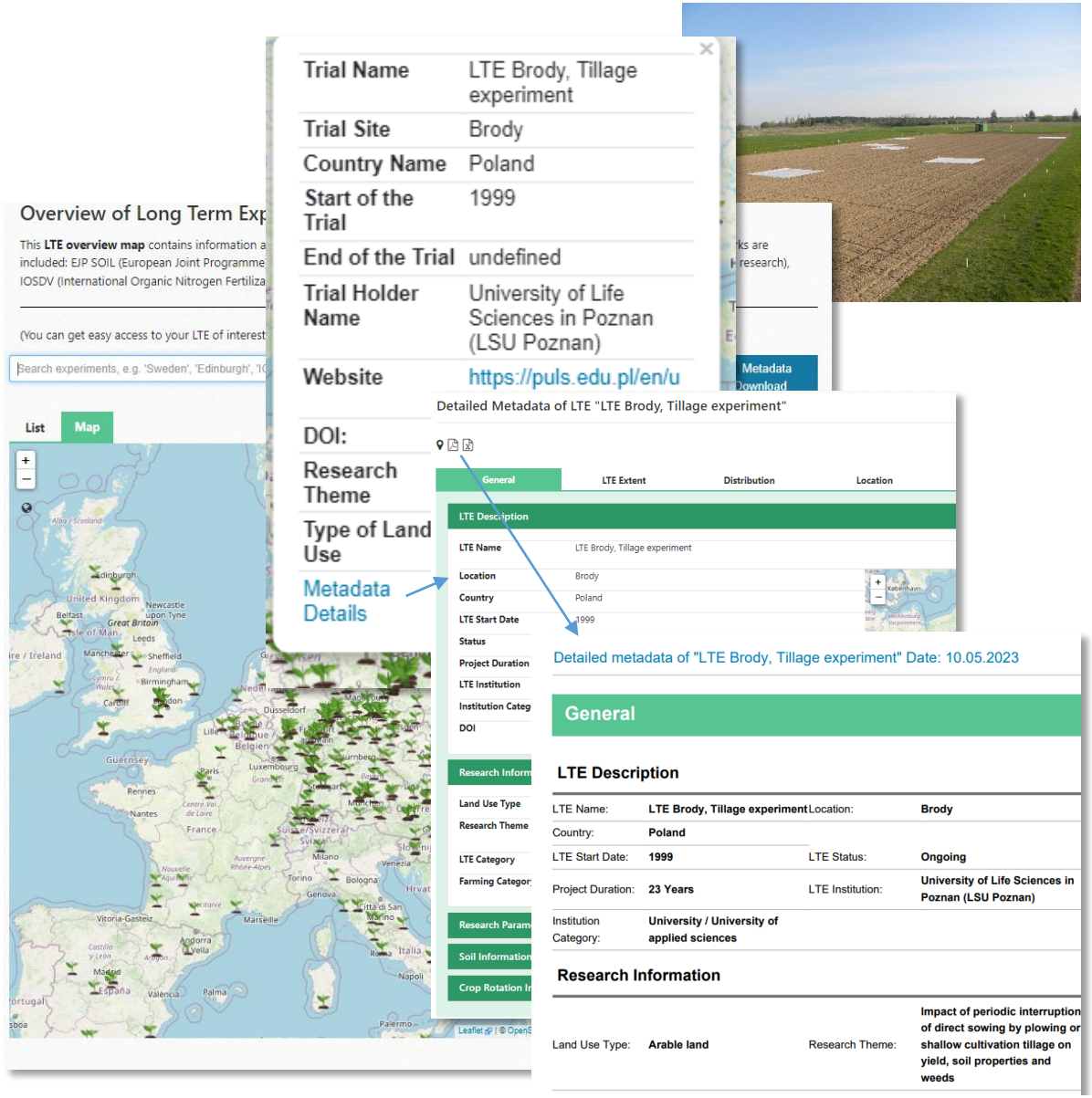
Find a browse (meta)data

LTE Overview map

- Agricultural long-term experiments (LTE)
- answer research questions of climate change, ecosystem services, nutrient cycles, agri. yield,…
- LTE definition: >20 years
- LTE data/contact often difficult to find / access

BonaRes product: map with metadata of ~ 570 LTE

- ✓ 30 countries
- ✓ metadata harmonization
- ✓ different display and filter functions
- ✓ open infrastructure
- ✓ international cooperation (GLTEN, EJP Soil)



lte.bonares.de