

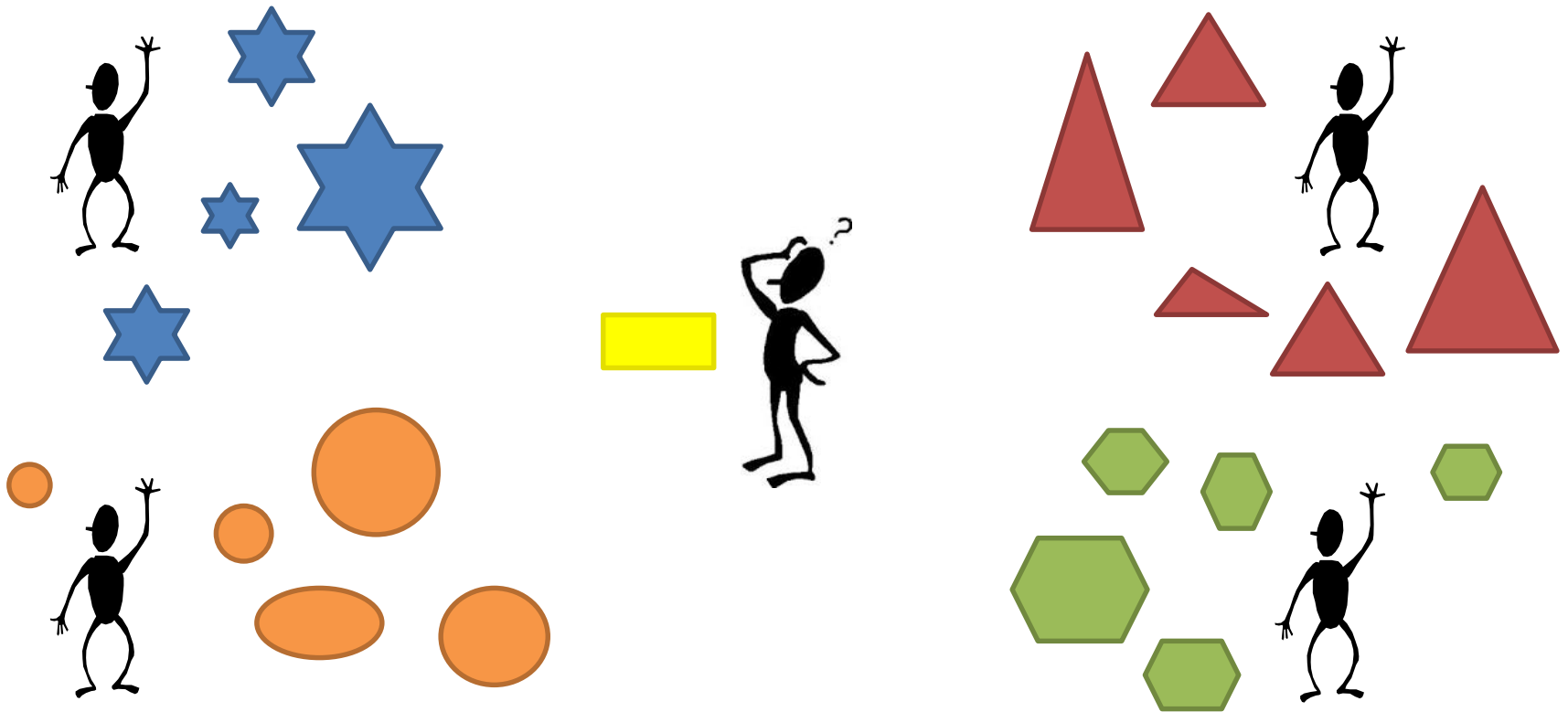
Methods of standardisation and annotation for plant phenotypic data

Paweł Krajewski, Hanna Ćwiek

Department of Biometry and Bioinformatics
Institute of Plant Genetics
Polish Academy of Science

The present situation

How to exchange and compare information about phenotyping experiments?



Standardisation is necessary!

Problem 1 – File formats

Materials and methods

Plant material

The study was performed on a population 99 line barley genotypes obtained by single seed descent (SSD) method and their parent genotypes: drought-tolerant Cam/B1 and drought-sensitive Maresi, received from the collection of the Institute of Plant Genetics (Poznan, Poland). Cam/B1 genotype (Cam/B1//CI008887/CI05761) is an advanced breeding line from Syria (ICARDA International Centre for Agricultural Research in the Dry Areas, Aleppo), and Maresi is a German (DDR) genotype, obtained from the Gene Bank in Prague (Czech Republic).

After sterilization, seeds were sown in pots with a capacity of 9 dm³ and filled with a mixture of soil and sand (3.5 : 1 v/v) at a 16-h photoperiod, an irradiance of 500 μmol(photon) m⁻² s⁻¹ (provided by high pressure sodium lamps, 400 W; Philips SON-T AGRO, Brussels, Belgium), and in 50% air humidity. Initially, 20 seedlings were placed in pots and after germination the number of plants was reduced to 10. During germination (4 days), a constant temperature of 25 °C was maintained. After emergence during the next 10 days, the temperature was 5 °C (day/night) and later, the temperature was 25/16 °C. Seedlings were watered and fertilized with a half-strength Hoagland nutrient solution (Hoagland and Arnon, 1938). Soil water content was determined by monitoring the weight (based on the water retention curve, ryc.1) and was stabilized to 8% water content of the dry weight of the soil (i.e. 3.2 pF). Soil drought (3.65%, i.e. 4.0 pF) was used after the appearance of the 4th leaf (16 day after emergence) and it was continued for 10 days. Plants grown in pots with 11 % water content were used as the control.

Measurements

The measurements were performed on the third leaf of the seedlings after 10 days of drought. The research concerned: (a) water relations of seedlings and integrity of cell membranes, (b) a fast chlorophyll *a* fluorescence kinetics - JIP test, (c) gas exchange, and (d) photochemical quenching and non-photochemical components of quenching (qE, qT, qI) under high light intensity conditions.

According to Bouslama and Schapaugh (1984), stress index (SI) were calculated for some of the parameters measured as follows: $SI(\%) = (X2/X1) \cdot 100\%$, where X2 and X1 represent the mean values of the parameters measured under drought stress and control.

Water relations of seedlings and the integrity of the cell membrane

Water relations were characterized by measurements of RWC and water content in the leaves. RWC was determined according to Barrs (1968): $RWC = (FW - DW) \times (TW - DW)^{-1} \cdot 100\%$; where FW is fresh weight, DW is dry weight and TW is turgid weight. To measure TW, leaf

Free text description – e.g. *.doc*, *.txt* file, article

Problem 1 – File formats

Experimentation											
Année	Surface parcelle semée	Surface parcelle récoltée	nb de rangs	ecartement des rangs	Panel de variétés	Nb variété	Irrigation	Azote	Fongicide	Répétition	
2011	24 m ²	10 m ²			2/3 Biotech partielle	98	IRR	X	F Opti	2	
2011	24 m ²	10 m ²			Biotech	156	SEC	X	F Opti	2	
2011											
2011											
Agricultural practices											
ID protocole	Opération	Date	Produit	Dose	Unité	modalité					
11GRE	Semis	2010-10-29	grains traités	250	gr/m ²	toutes					
11GRE	Fongicide					irrigué					
11GRE	Engrais					sec					
11GRE	Irrigation					tout sauf sec					
11GRE	...										

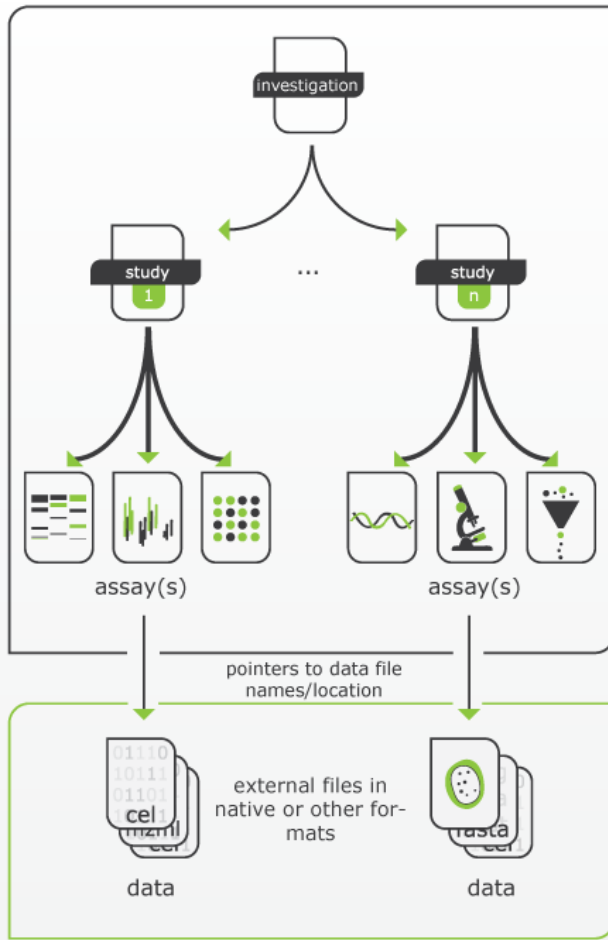
Tables, spraed sheets – e.g. .x/s file

Problem 1 – File formats

```
<?xml version = '1.0' encoding = 'ISO-8859-1' ?>
<!-- <!DOCTYPE XMI SYSTEM 'UML13.dtd' > -->
<XMI xmi.version = '1.0' timestamp = 'Wed Sep 04 10:51:34 2002' >
  <XMI.header>
    <XMI.documentation>
      <XMI.exporter>Unisys.JCR.1</XMI.exporter>
      <XMI.exporterVersion>1.3.2</XMI.exporterVersion>
    </XMI.documentation>
    <XMI.metamodel xmi.name = 'UML' xmi.version = '1.3' />
  </XMI.header>
  <XMI.content>
    <!-- ===== MAGE-OM [Model] ===== -->
    <Model_Management.Model xmi.id = 'G.0' >
      <Foundation.Core.ModelElement.name>MAGE-OM</Foundation.Core.ModelElement.name>
      <Foundation.Core.ModelElement.visibility xmi.value = "public"/>
      <Foundation.Core.ModelElement.isSpecification xmi.value = "false"/>
      <Foundation.Core.GeneralizableElement.isRoot xmi.value = "false"/>
      <Foundation.Core.GeneralizableElement.isLeaf xmi.value = "false"/>
      <Foundation.Core.GeneralizableElement.isAbstract xmi.value = "false"/>
      <Foundation.Core.Namespace.ownedElement>
        <!-- ===== MAGE-OM::ArrayDesign [Package] ===== -->
        <Model_Management.Package xmi.id = 'S.10' >
          <Foundation.Core.ModelElement.name>ArrayDesign</Foundation.Core.ModelElement.name>
          <Foundation.Core.ModelElement.visibility xmi.value = "public"/>
          <Foundation.Core.ModelElement.isSpecification xmi.value = "false"/>
          <Foundation.Core.GeneralizableElement.isRoot xmi.value = "false"/>
        </Model_Management.Package>
      </Foundation.Core.Namespace.ownedElement>
    </Model_Management.Model>
  </XMI.content>
</XMI>
```

Machine readable formats – e.g. *.xml* file

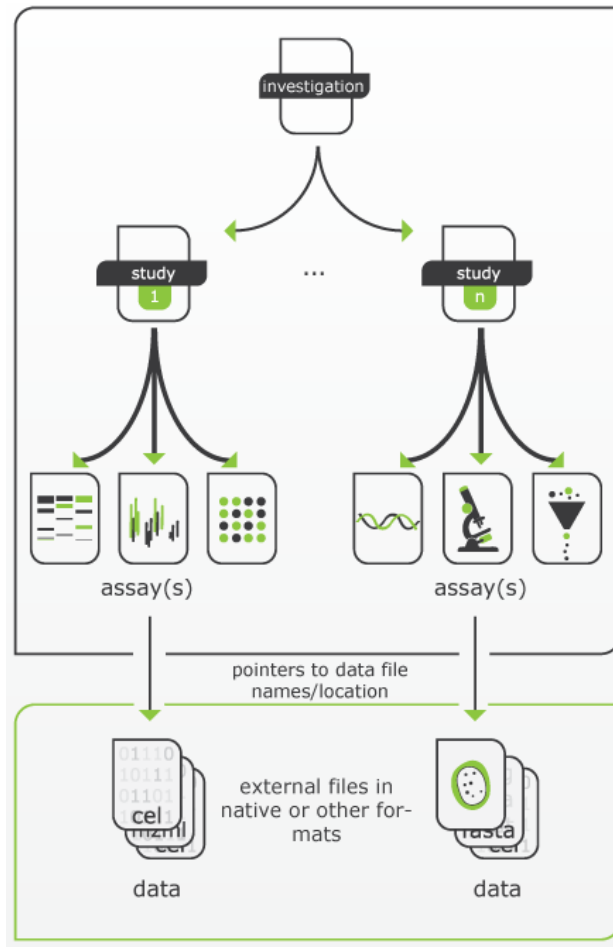
Solution 1 – Standard file format



ISA-TAB file

- Universal schema
- Already accepted in other domains, e.g. metabolomics
- Human- and machine-readable
- Hierarchy of text files

Structure of ISA-TAB format



Investigation

General information about goals, researchers, methodology, publications

Study

Information about biosource: species, variety, treatment

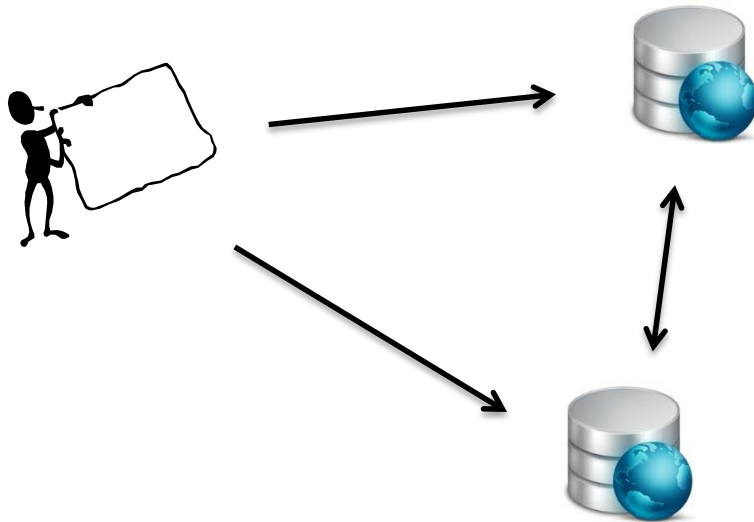
Assay

Information about measurements: technology, parameters, protocols, replications, methods of processing of data

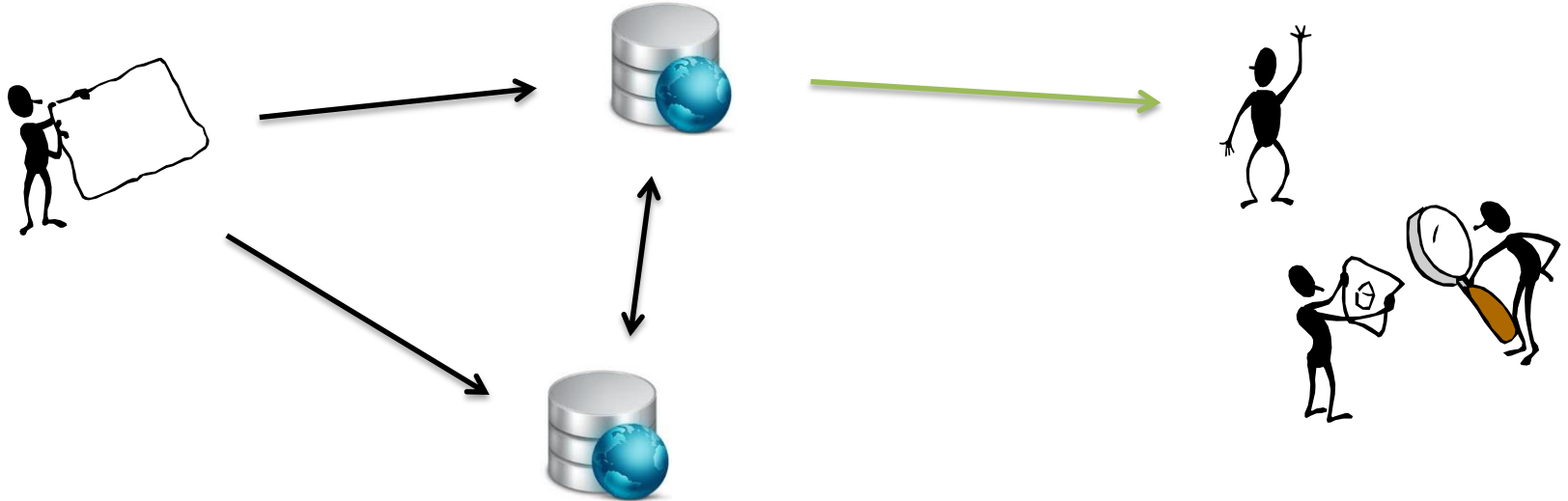
Data

Measurements: raw data, processed data, traits descriptions

Is a common data format enough?

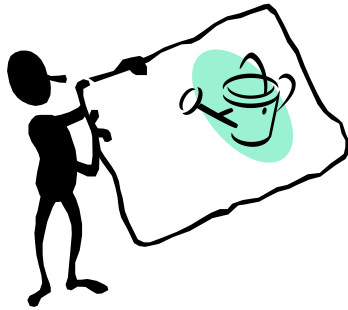


Is a common data format enough?



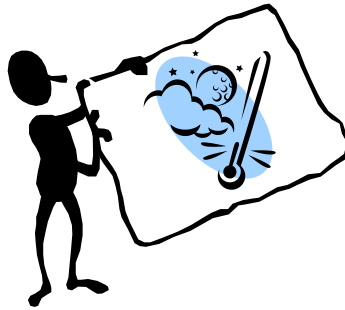
Some information is missing!

Problem 2 – Necessary information



Variable **water** regimen

„Other factors for all plant were **the same**”



Variable **temperature**

„Other factors for all plant were **the same**”



Different plant **varieties**

„Other factors for all plants were **the same**”

All factors that could influence results must be reported

Solution 2 – MI standard

Minimum Information About ...

- minimum amount of metadata for specific aim
- accepted in biological research, e.g.
 - ... Microarray Experiment (MIAME)
 - ... Proteomic Experiment (MIAPE)
 - ... Genome Sequences (MIGS)
 - ... **Plant Phenotyping Experiment (MIAPPE)**
- checklist



Solution 2 – MI standard

- Study
- Environment
- Biosource, treatments
- Experimental design
- Sample
collection, processing, management
- Phenotypic traits
type, measurement protocol, processing protocol,
scale, units



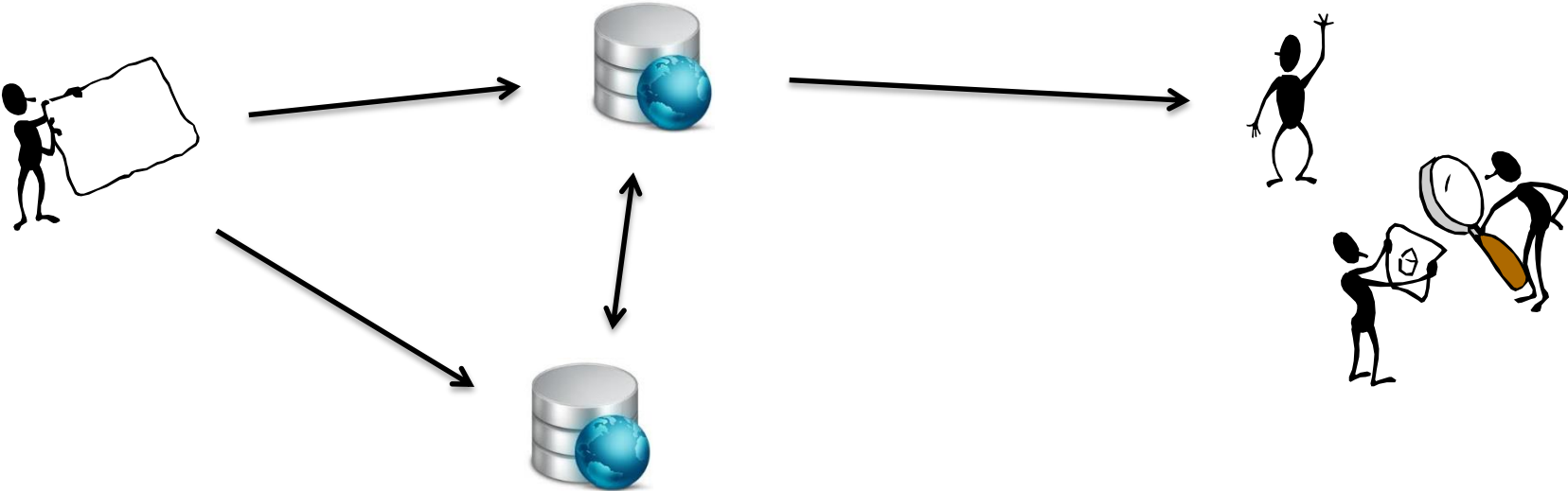
Solution 2 – MI standard

- Study
- Environment
- Biosource, treatments
- Experimental design
- Sample
collection, processing, management
- Phenotypic traits
type, measurement protocol, processing protocol,
scale, units

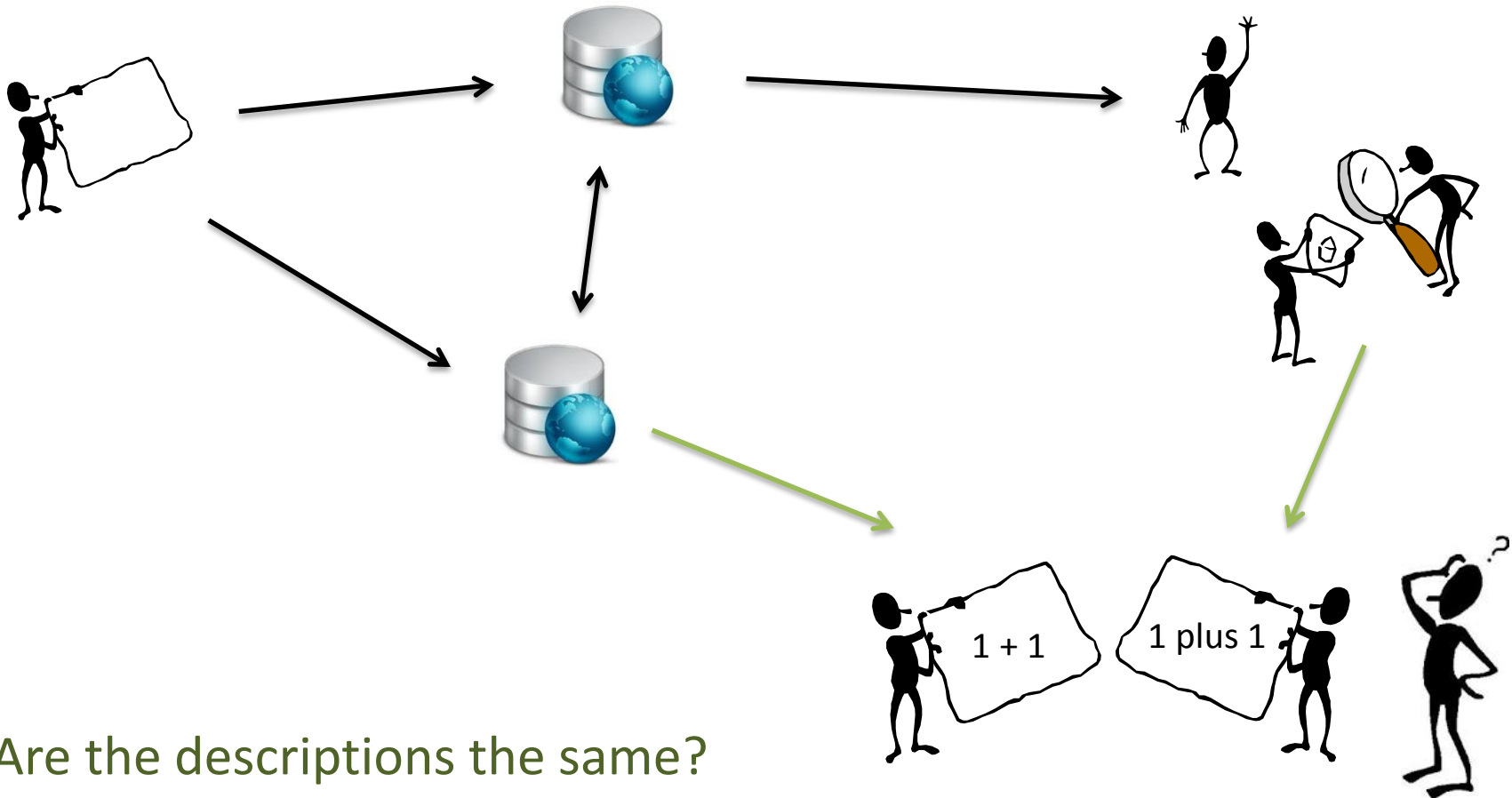


Is mentioning all information enough?

Is mentioning all the information enough?

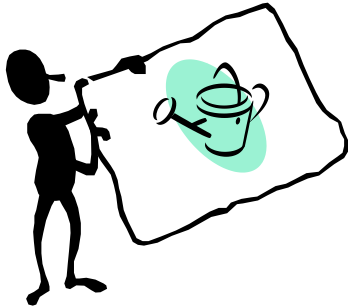


Is mentioning all the information enough?



Are the descriptions the same?

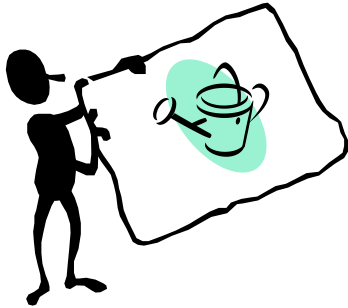
Problem 3 – Clarity of concepts



- „Water content”
- „Relative water content”
- „Leaf water content”
- „Content of water in leaves”
- „Moisture”
- „Tissue moisture”
- „Zawartość wody”
- „Wilgotność względna”
- ...

Different names, the same concepts
Different concepts, the same name

Problem 3 – Clarity of concepts



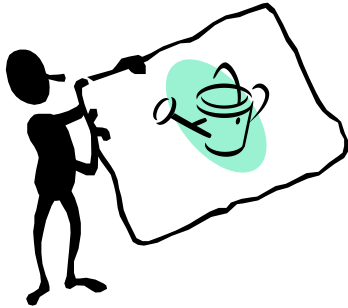
- „Water content”
- „Relative water content”
- „Leaf water content”
- „Content of water in leaves”
- „Moisture”
- „Tissue moisture”
- „Zawartość wody”
- „Wilgotność względna”
- ...

Different names, the same concepts
Different concepts, the same name

Observable traits

- what is measured? plant part, property
- how measured? device, parameters, sample size
- how processed? programs, algorithms
- how expressed? Units

Problem 3 – Clarity of concepts



- „Water content”
- „Relative water content”
- „Leaf water content”
- „Content of water in leaves”
- „Moisture”
- „Tissue moisture”
- „Zawartość wody”
- „Wilgotność względna”
- ...

Different names, the same concepts
Different concepts, the same name

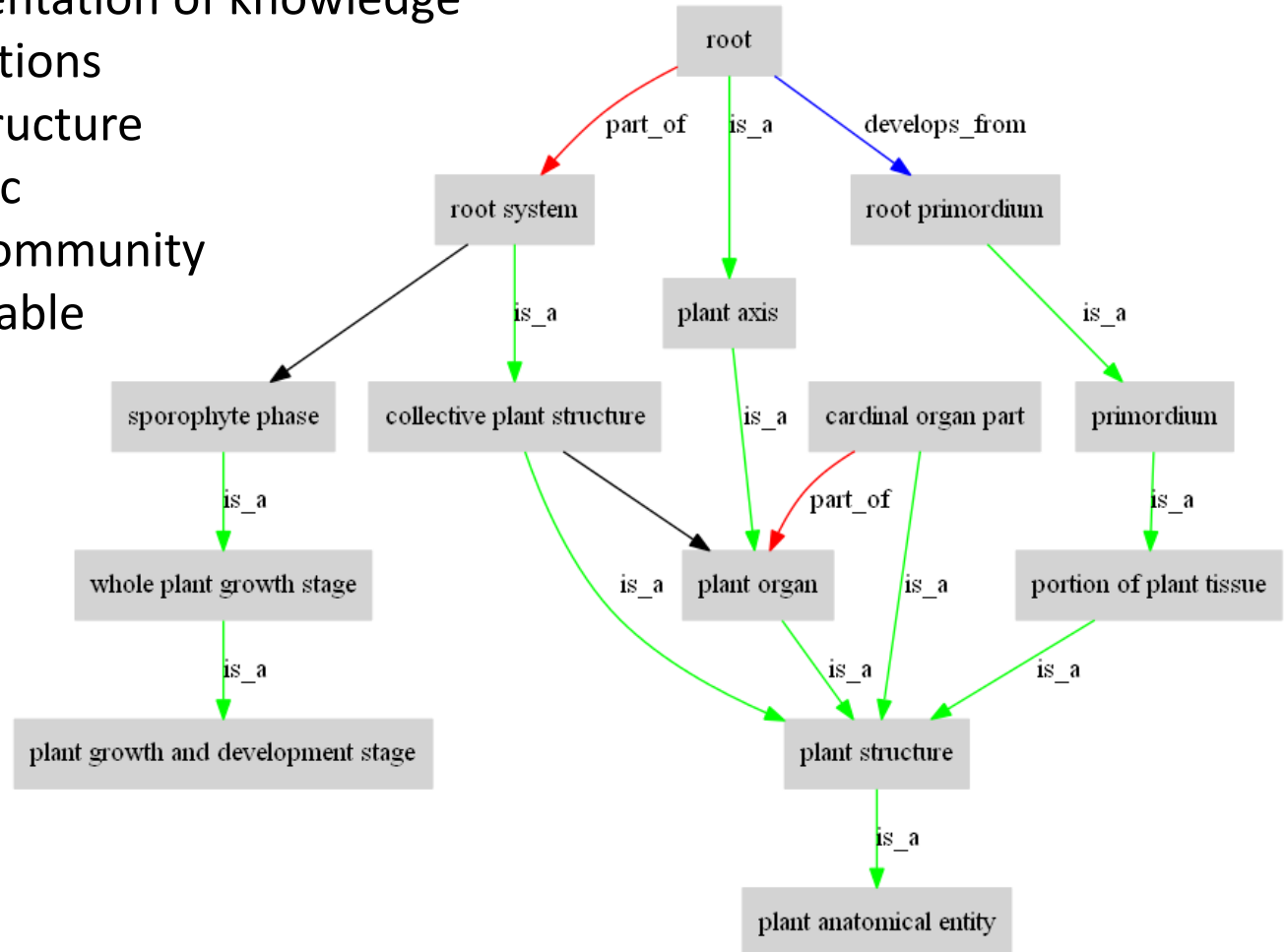
Observable traits

- what is measured? plant part, property
- how measured? device, parameters, sample size
- how processed? programs, algorithms
- how expressed? Units

How to assure clarity of description?

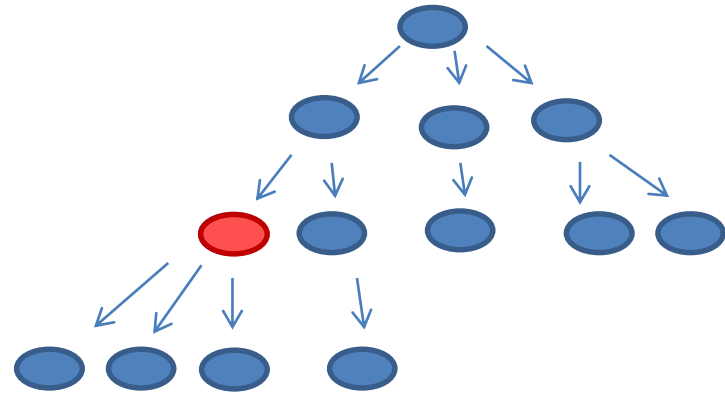
Solution 3 - Ontologies

- Formal representation of knowledge
- Terms and relations
- Hierarchical structure
- Domain specific
- Created by a community
- Publically available



Solution 3 - Ontologies

- „Water content”
XO:00315
- „Relative water content”
- „Moisture”
- „Tissue moisture”
- „Zawartość wody”
- „Wilgotność względna”

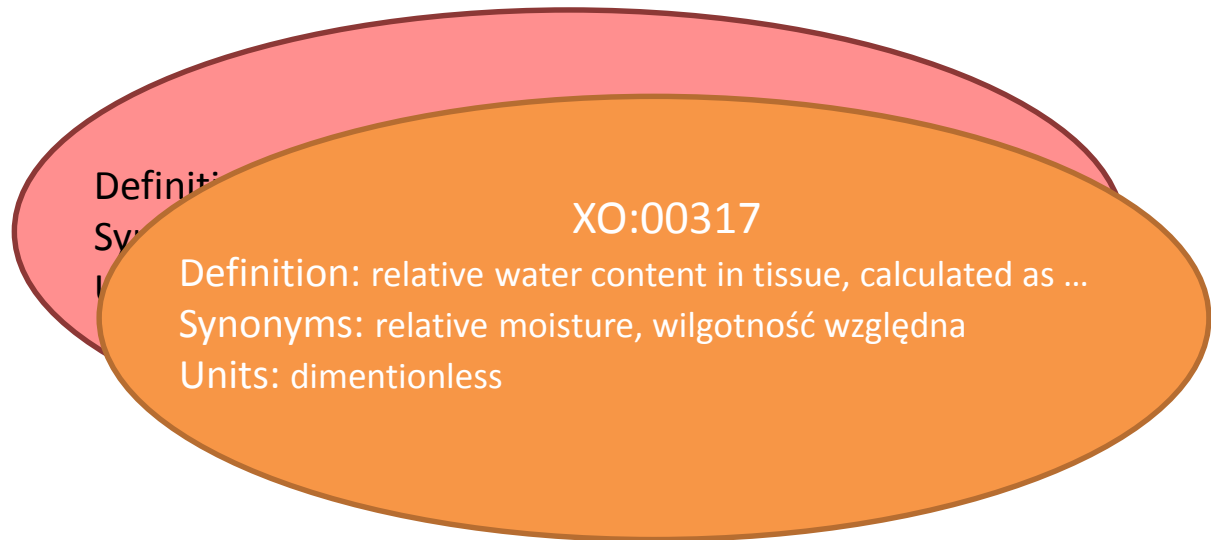
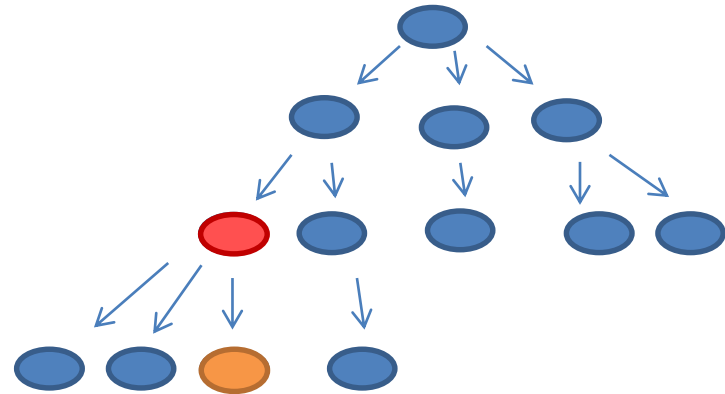


XO:00315

Definition: water content in tissue measured as % of dry mass
Synonyms: moisture, water content, zawartość wody
Units: %

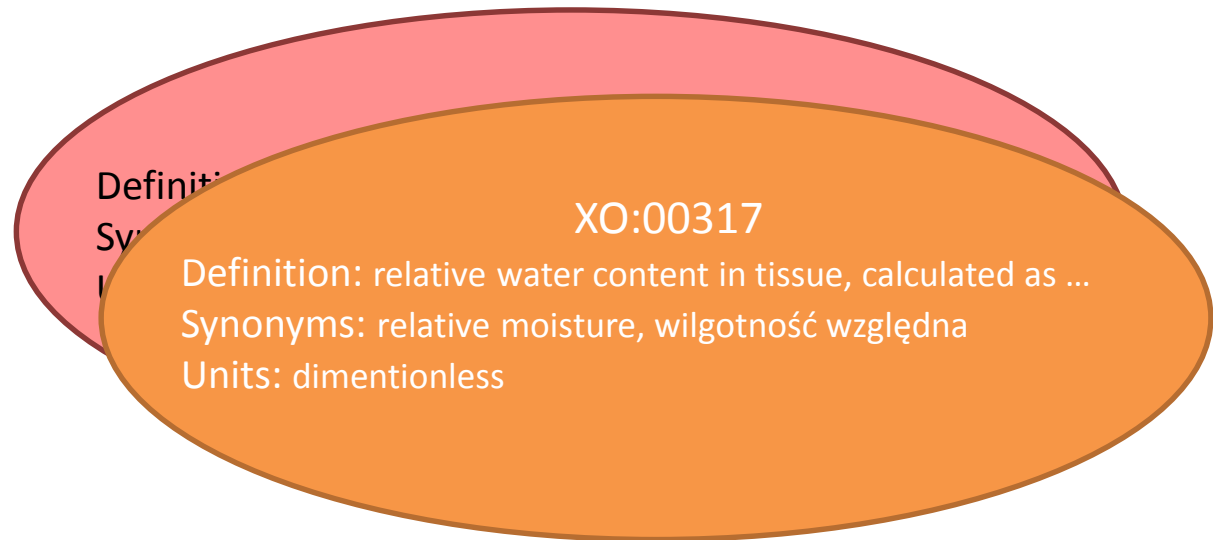
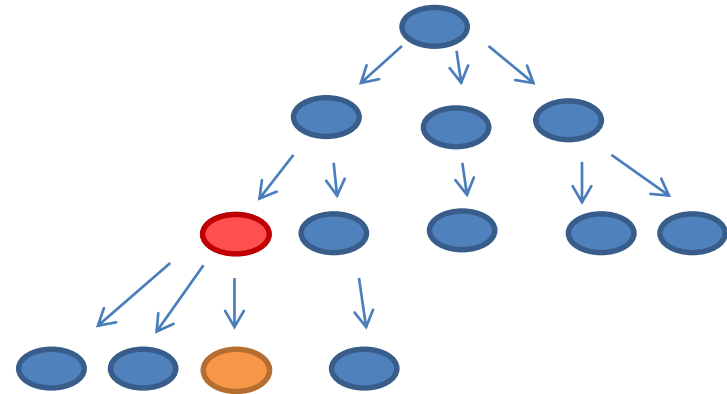
Solution 3 - Ontologies

- „Water content”
XO:00315
- „Relative water content”
XO:00317
- „Moisture”
- „Tissue moisture”
- „Zawartość wody”
- „Wilgotność względna”

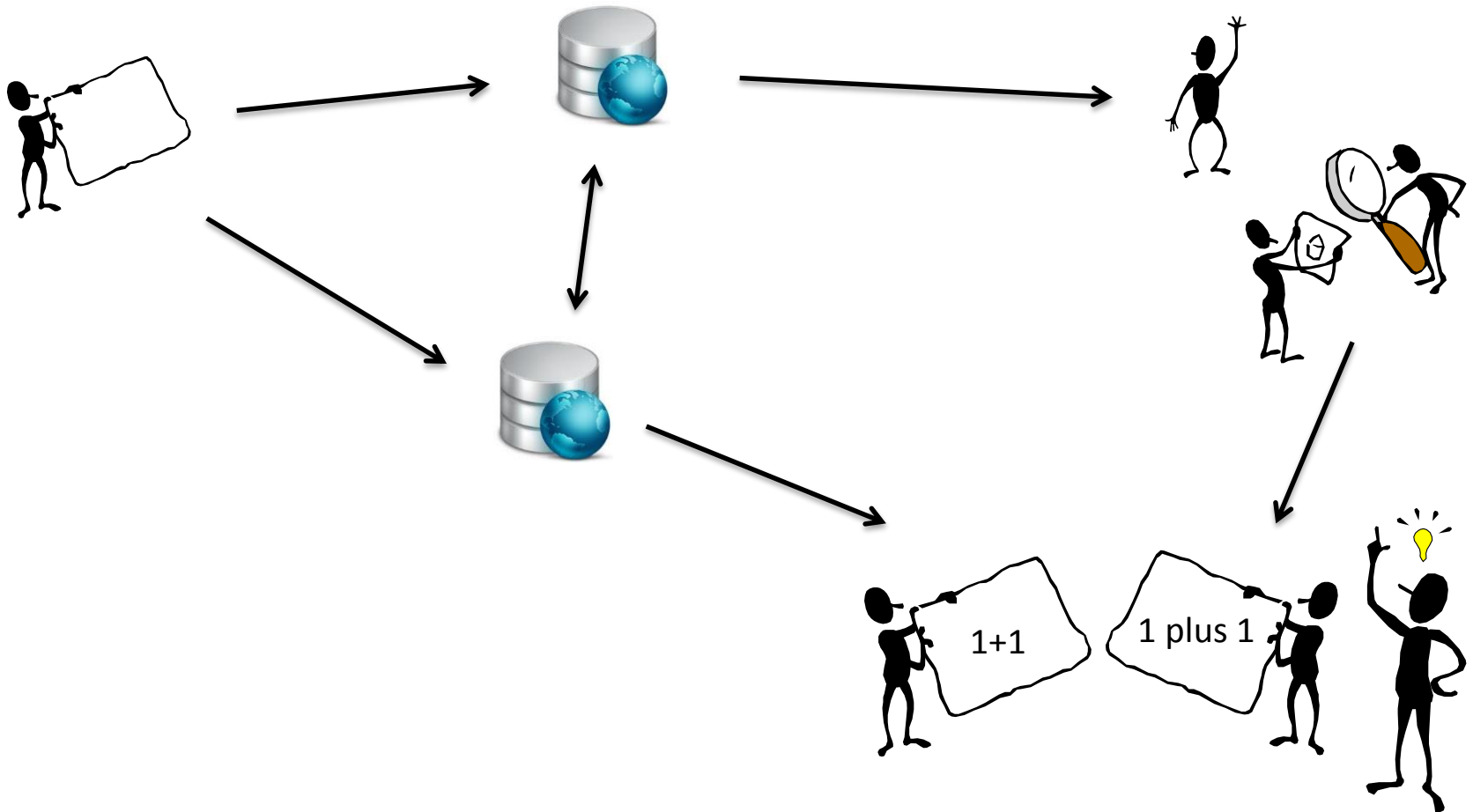


Solution 3 - Ontologies

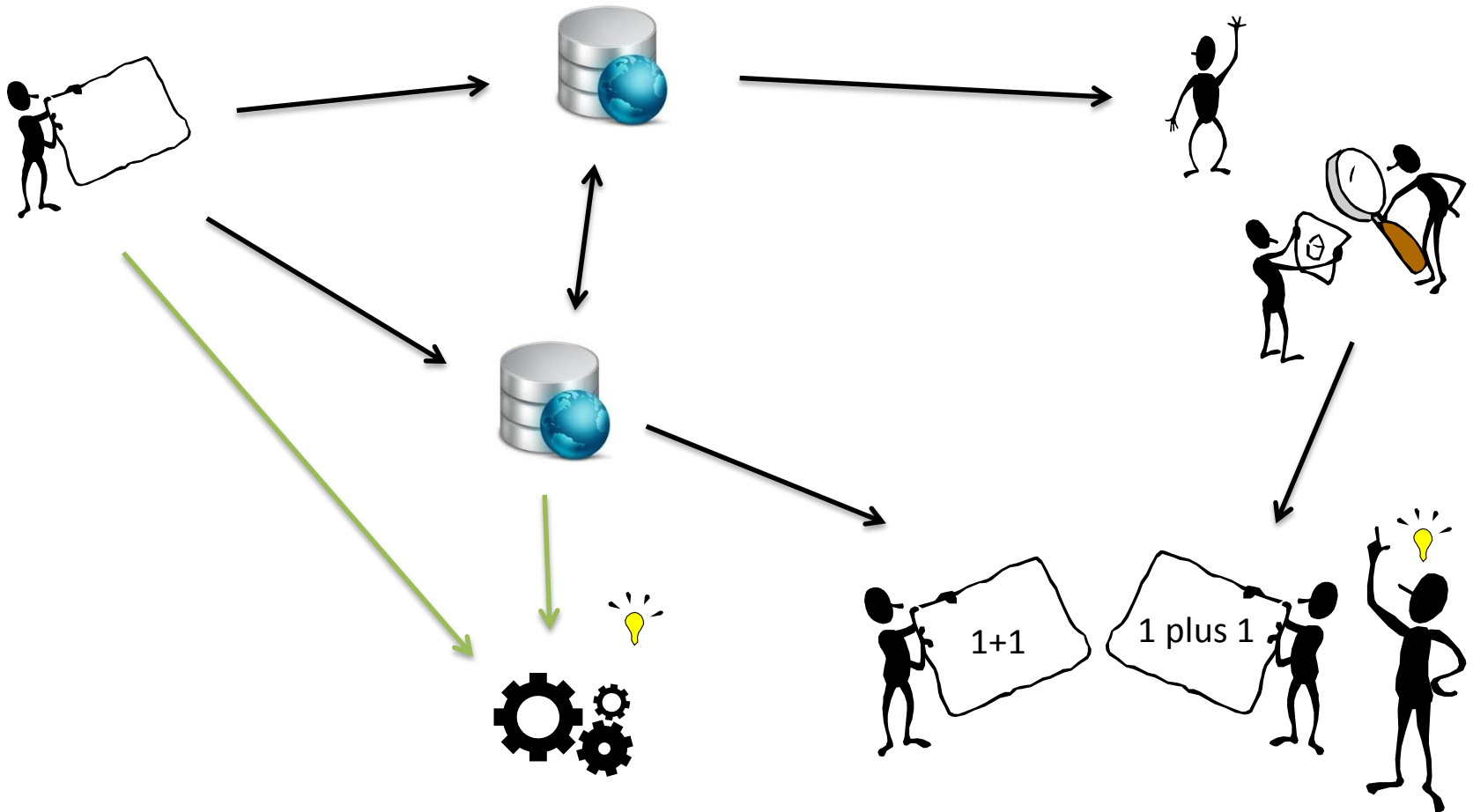
- „Water content”
XO:00315
- „Relative water content”
XO:00317
- „Moisture”
XO:00315
- „Tissue moisture”
XO:00315
- „Zawartość wody”
XO:00315
- „Wilgotność względna”
XO:00317



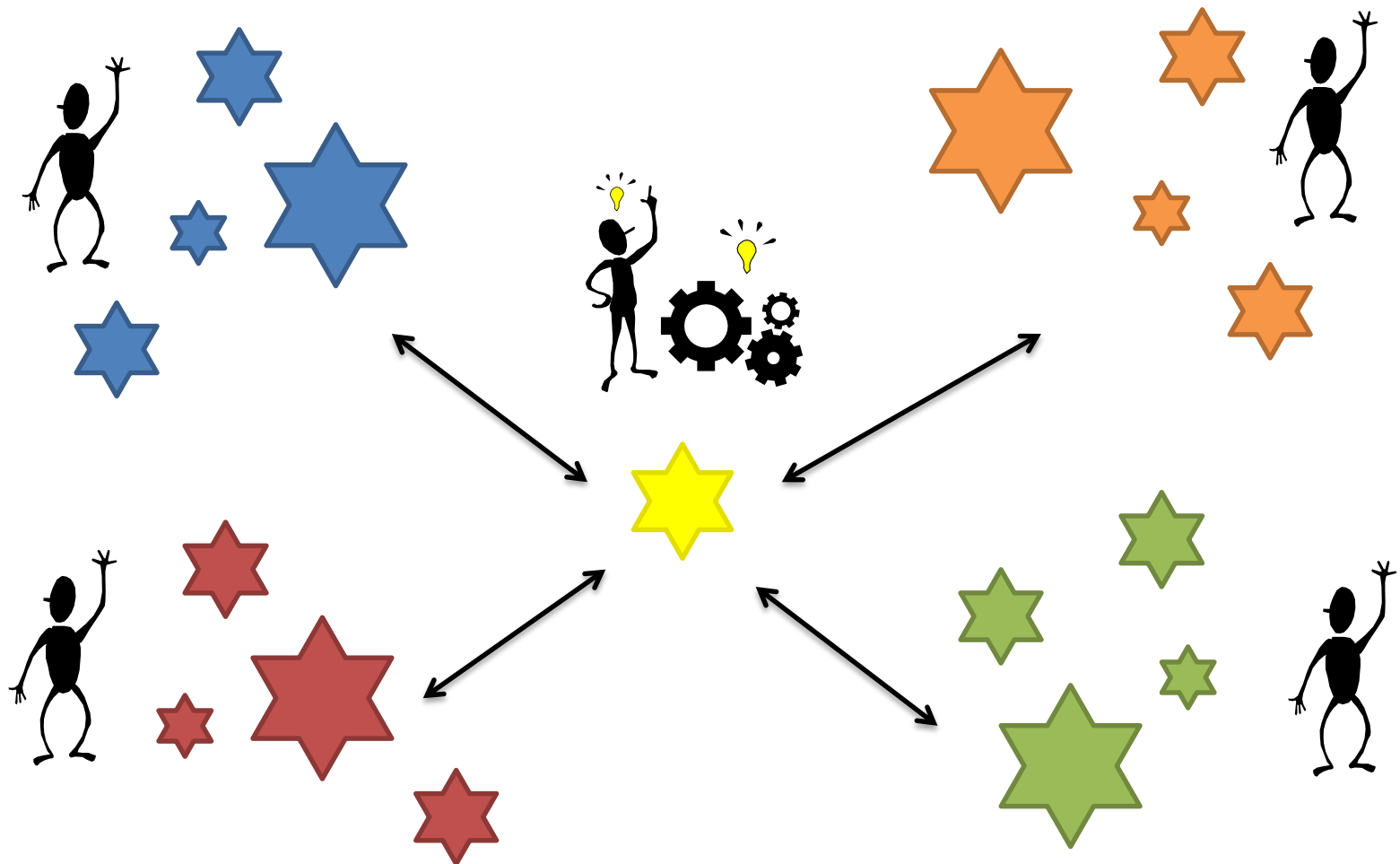
Understandable standards



Understandable standards



The ideal situation



Hands-on

Consider an experiment

- Aim: study tolerance of barley varieties to drought
- Experimental design: complete blocks, 5 replications
- Measured traits – phenotypic:
 - stem length
 - stem colour
 - stem diameter
- Presented as simplified ISA-TAB structure:
<http://tinyurl.com/workshopPoznan> → IPG → excercise.xlsx

Hands-on

Task: Annotate the experiment

- Add missing information (marked as ‘?’) by annotating corresponding free text with ontology terms
- Use the ontology browsing tools to search for matching terms (see next page)
- Some of the suggested ontologies are specified at the beginning of investigation file (section ‘Ontology sources reference’)

Resources

Ontology browsers:

- <http://www.ebi.ac.uk/ontology-lookup/>
- <http://bioportal.bioontology.org/>
- http://gramene.org/plant_ontology/
- <http://www.cropontology.org/>

Other links

- Minimum Information standards

<http://mibbi.sourceforge.net/portal.shtml>

- ISA-TAB format

<http://isatab.sourceforge.net/format.html>

Thank you for your attention