Standaryzacja i publikacja danych w badaniach nad roślinami

Paweł Krajewski

Zakład Biometrii i Bioinformatyki
Instytut Genetyki Roślin PAN
w Poznaniu

pkra@igr.poznan.pl

www.transplantdb.eu

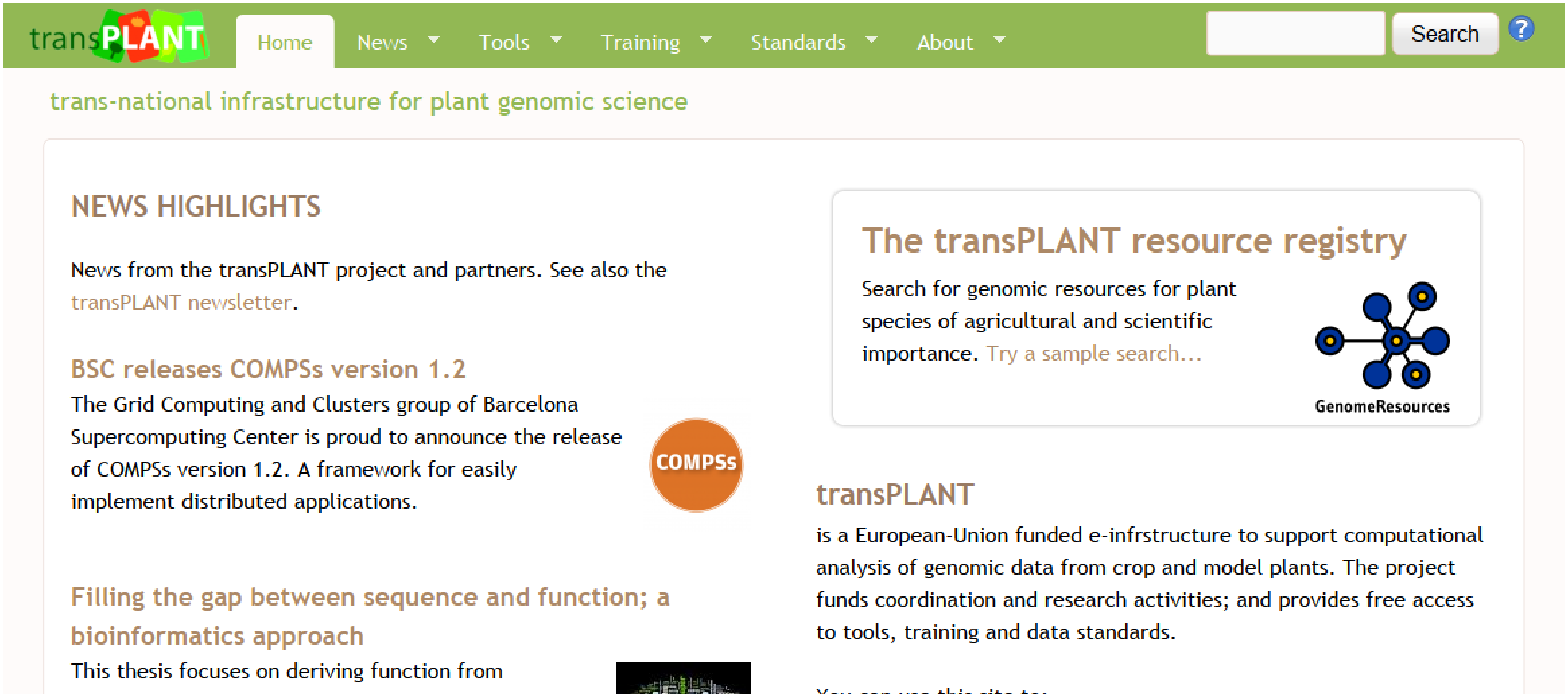
The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Projekt transPLANT

- 4-letni projekt (2011-2015) EU FP7 (DG CONNECT) koordynowany przez EMBL-European Bioinformatic Institute, Hinxton, UK.
- Cel: zaprojektować, rozmieścić i zarządzać infrastrukturą (oprogramowaniem) ważną dla przyszłych potrzeb badaczy roślin.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Projekt transPLANT



trans-national infrastructure for plant genomic science

NEWS HIGHLIGHTS

News from the transPLANT project and partners. See also the transPLANT newsletter.

BSC releases COMPSs version 1.2

The Grid Computing and Clusters group of Barcelona Supercomputing Center is proud to announce the release of COMPSs version 1.2. A framework for easily implement distributed applications.

Filling the gap between sequence and function; a bioinformatics approach

This thesis focuses on deriving function from

The transPLANT resource registry

Search for genomic resources for plant species of agricultural and scientific importance. Try a sample search…

transPLANT

is a European-Union funded e-infrstructure to support computational analysis of genomic data from crop and model plants. The project funds coordination and research activities; and provides free access to tools, training and data standards.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane biologiczne

Typy danych

- experiment, sample, file identification metadata,
- environmental data,
- accession, pedigree, passport data,
- genomic sequence, sequence annotation and variation data,
- Next Generation Sequencing raw and derived data,
- QTL and GWAS data,
- transcriptomic, proteomic, metabolomic („omics”) data,
- traditional and high-throughput (image) phenotyping data,
- genetic marker data,
- physical, genetic, other map data,
- breeding, food industry and market-related data,
- literature data.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dlaczego standaryzować?

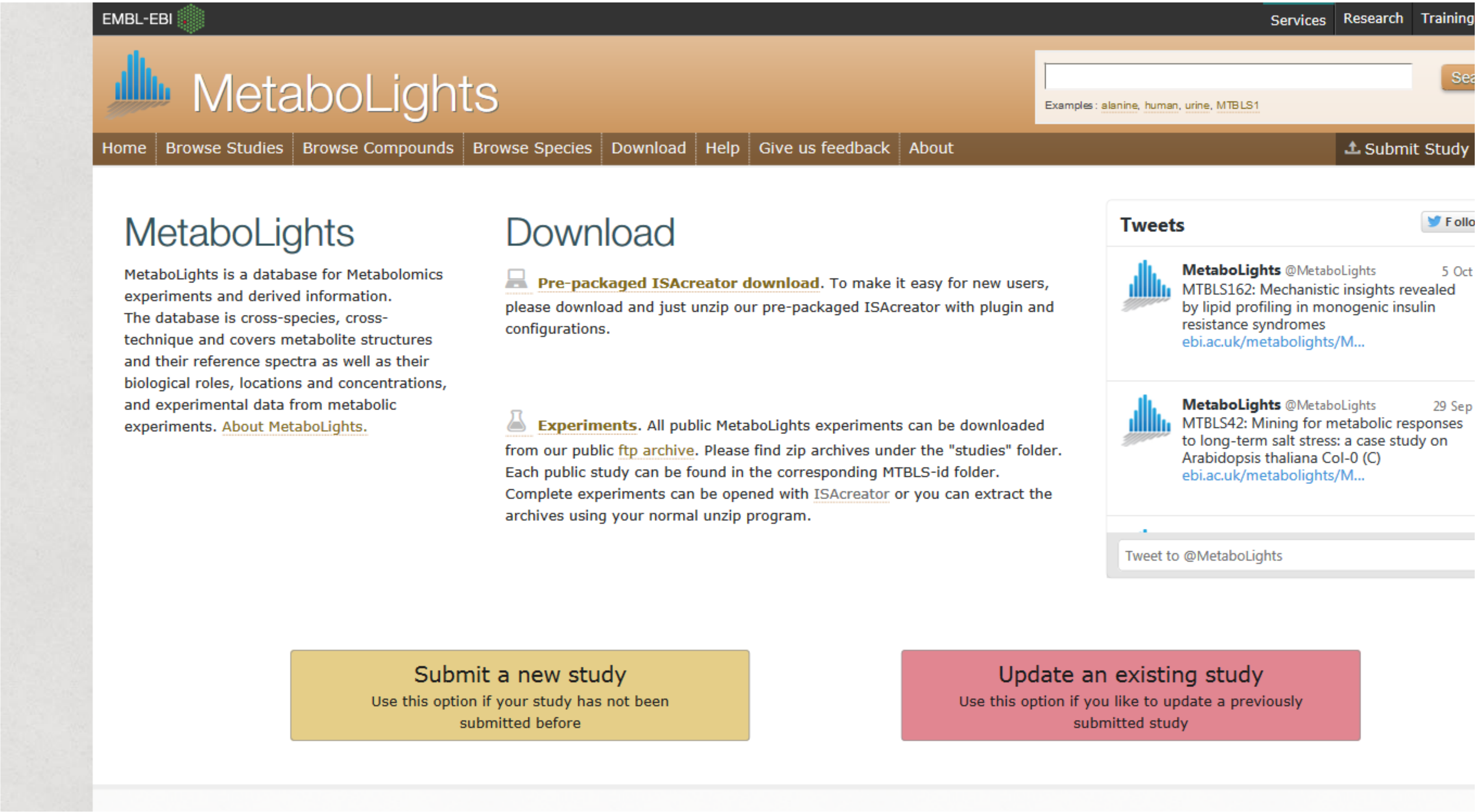
- Aby móc zapewnić otwartość danych przy zachowaniu opisu ich znaczenia i struktury

- Aby ułatwić (= obniżyć koszt) przechowywanie, publikację, wymianę między bazami danych, ...

- Wyszukiwanie doświadczeń (poprzez metadane lub dane)
- Porównywanie doświadczeń
- Integracja danych
- Analiza statystyczna sterowana metadanymi i ich adnotacją

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane metabolomiczne



MetaboLights

MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from metabolic experiments. About MetaboLights.

Download

Pre-packaged ISAcreator download. To make it easy for new users, please download and just unzip our pre-packaged ISAcreator with plugin and configurations.

Experiments. All public MetaboLights experiments can be downloaded from our public ftp archive. Please find zip archives under the "studies" folder. Each public study can be found in the corresponding MTBLS-id folder. Complete experiments can be opened with ISAcreator or you can extract the archives using your normal unzip program.

Tweets

MetaboLights @MetaboLights 5 Oct
MTBLS162: Mechanistic insights revealed by lipid profiling in monogenic insulin resistance syndromes
ebi.ac.uk/metabolights/M...

MetaboLights @MetaboLights 29 Sep
MTBLS42: Mining for metabolic responses to long-term salt stress: a case study on Arabidopsis thaliana Col-0 (C)
ebi.ac.uk/metabolights/M...

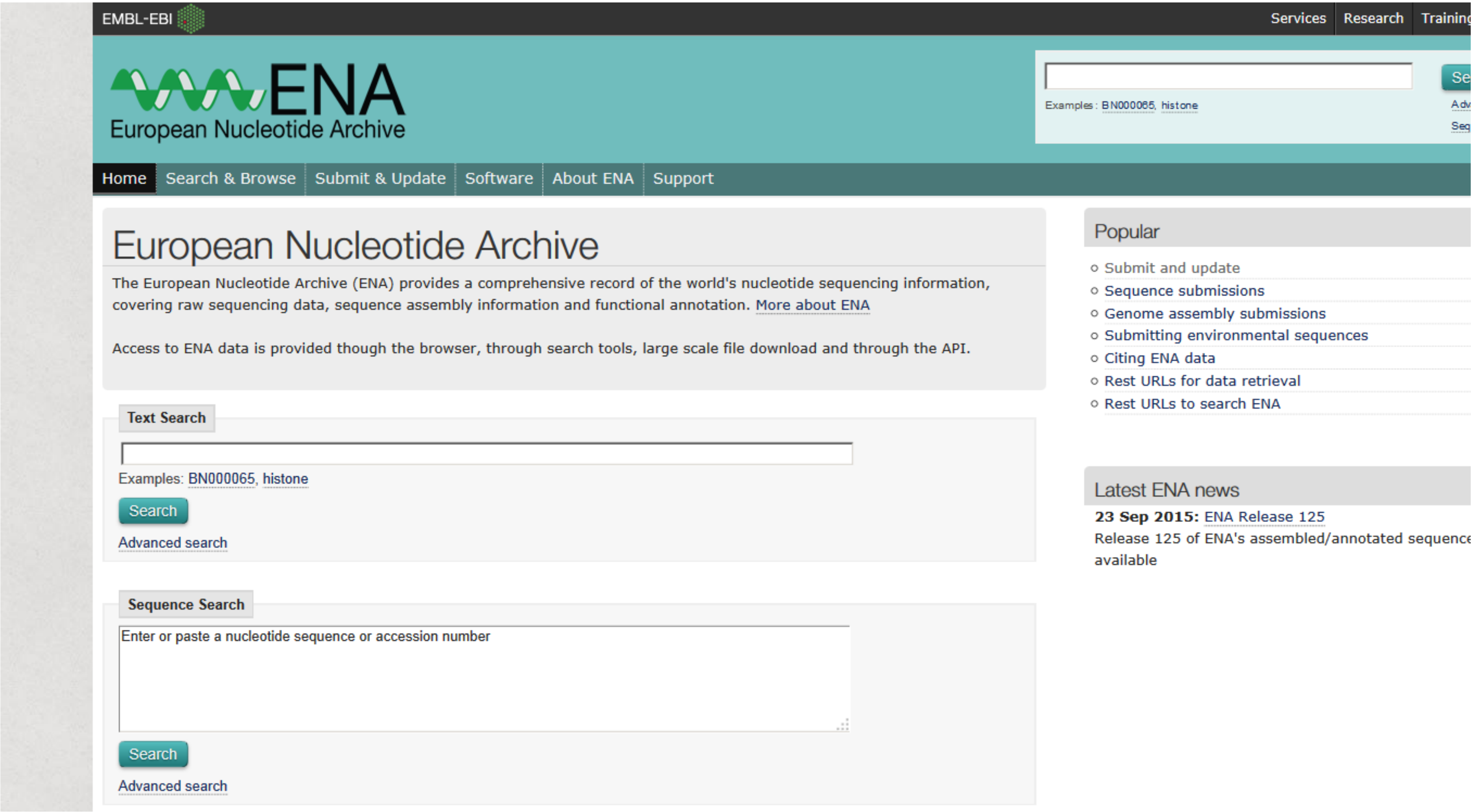
Tweet to @MetaboLights

Submit a new study
Use this option if your study has not been submitted before

Update an existing study
Use this option if you like to update a previously submitted study

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Sekwencje nukleotydów



European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. More about ENA

Access to ENA data is provided though the browser, through search tools, large scale file download and through the API.

Text Search

Examples: BN000065, histone

Search

Advanced search

Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

Advanced search

Popular

- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- Rest URLs for data retrieval
- Rest URLs to search ENA

Latest ENA news

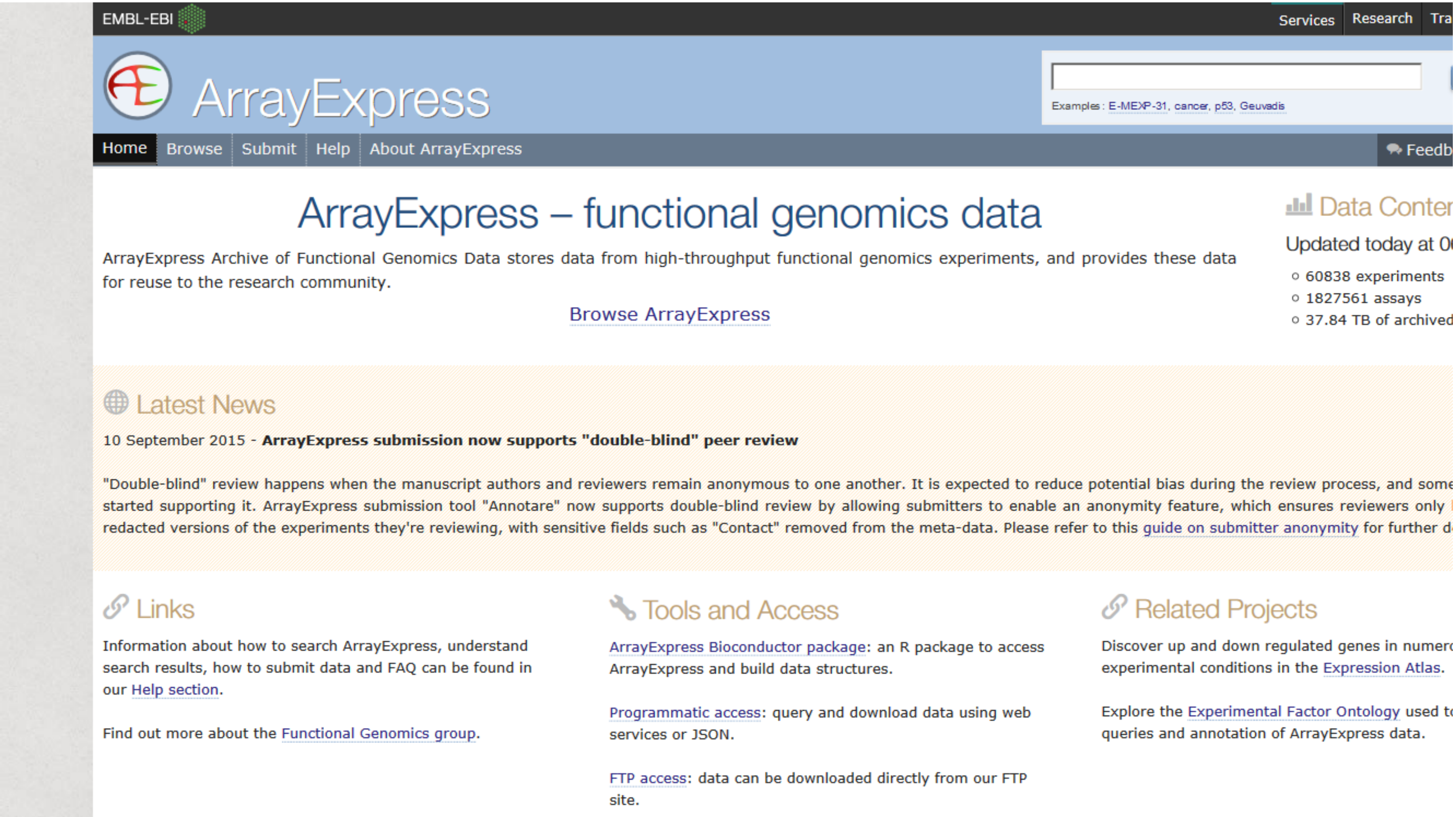
23 Sep 2015: ENA Release 125

Release 125 of ENA's assembled/annotated sequence

available

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane o ekspresji genów



ArrayExpress – functional genomics data

ArrayExpress Archive of Functional Genomics Data stores data from high-throughput functional genomics experiments, and provides these data for reuse to the research community.

Browse ArrayExpress

Data Conter

Updated today at 0

- 60838 experiments
- 1827561 assays
- 37.84 TB of archived

Latest News

10 September 2015 - **ArrayExpress submission now supports "double-blind" peer review**

"Double-blind" review happens when the manuscript authors and reviewers remain anonymous to one another. It is expected to reduce potential bias during the review process, and some started supporting it. ArrayExpress submission tool "Annotare" now supports double-blind review by allowing submitters to enable an anonymity feature, which ensures reviewers only redacted versions of the experiments they're reviewing, with sensitive fields such as "Contact" removed from the meta-data. Please refer to this guide on submitter anonymity for further d

Links

Information about how to search ArrayExpress, understand search results, how to submit data and FAQ can be found in our Help section.

Find out more about the Functional Genomics group.

Tools and Access

ArrayExpress Bioconductor package: an R package to access ArrayExpress and build data structures.

Programmatic access: query and download data using web services or JSON.

FTP access: data can be downloaded directly from our FTP site.

Related Projects

Discover up and down regulated genes in numer experimental conditions in the Expression Atlas.

Explore the Experimental Factor Ontology used t queries and annotation of ArrayExpress data.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane biologiczne

Typy danych

- experiment, sample, file identification metadata,
- environmental data,
- accession, pedigree, passport data,
- genomic sequence, sequence annotation and variation data,
- Next Generation Sequencing raw and derived data,
- QTL and GWAS data,
- transcriptomic, proteomic, metabolomic („omics”) data,
- **traditional and high-throughput (image) phenotyping data,**
- genetic marker data,
- physical, genetic, other map data,
- breeding, food industry and market-related data,
- literature data.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane fenotypowe

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Dane fenotypowe

- **Dane fenotypowe:** obserwacje cech morfologicznych, odporności, składników plonu, fizjologiczne, biochemiczne, ...

- **Charakterystyczne cechy:**
  - zależność od genotypu i środowiska,
  - duże spektrum cech, metod próbkowania, metod pomiarowych, jednostek.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.

Minimum Information about Plant Phenotypic Experiment (MIAPPE)

Doświadczenie

Atrybuty

Próby (jednostki)

s_ad_1	3	Dro1	Maresi	…	…	1.34	2.14	red	…

„Checklist”: lista atrybutów, które należy podać

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

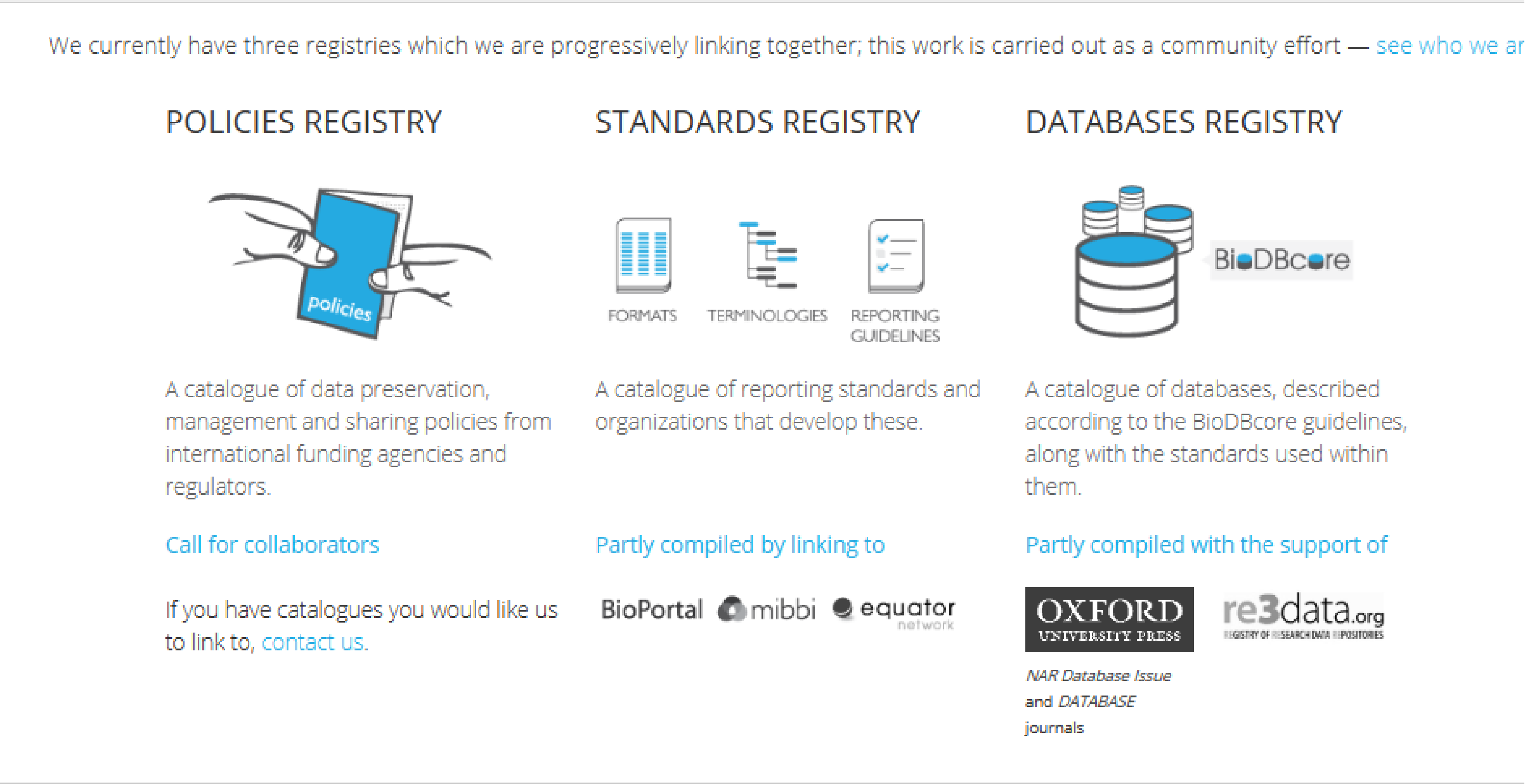
Minimum Information about Plant Phenotypic Experiment (MIAPPE)

Checklist section	Attributes
General metadata	Unique identifier, Title, Description
Timing and location	Start of experiment (date), Duration (days/months/years), Geographic location
Environment	Growth facility, Aerial conditions (air humidity, light intensity, temperature) Rooting conditions (medium, plot size, sowing density, pH) Nutrients Watering (range in water potential, irrigation type)
Biosource	Organism (species), Infraspecific name, Seed origin
Treatments	Treatment factors (name, levels)
Experimental design	Blocking factors (name, levels)
Sample collection, management	Plant body site (organ)
Variables	Phenotypic variables (name, method, scale) Environmental variables (name, method, scale)
Observations	Raw data, Processing protocol, Derived data

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.

Minimum Information about Plant Phenotypic Experiment (MIAPPE)

Biosharing



We currently have three registries which we are progressively linking together; this work is carried out as a community effort — see who we ar

POLICIES REGISTRY

A catalogue of data preservation, management and sharing policies from international funding agencies and regulators.

Call for collaborators

If you have catalogues you would like us to link to, contact us.

STANDARDS REGISTRY

A catalogue of reporting standards and organizations that develop these.

Partly compiled by linking to

DATABASES REGISTRY

A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them.

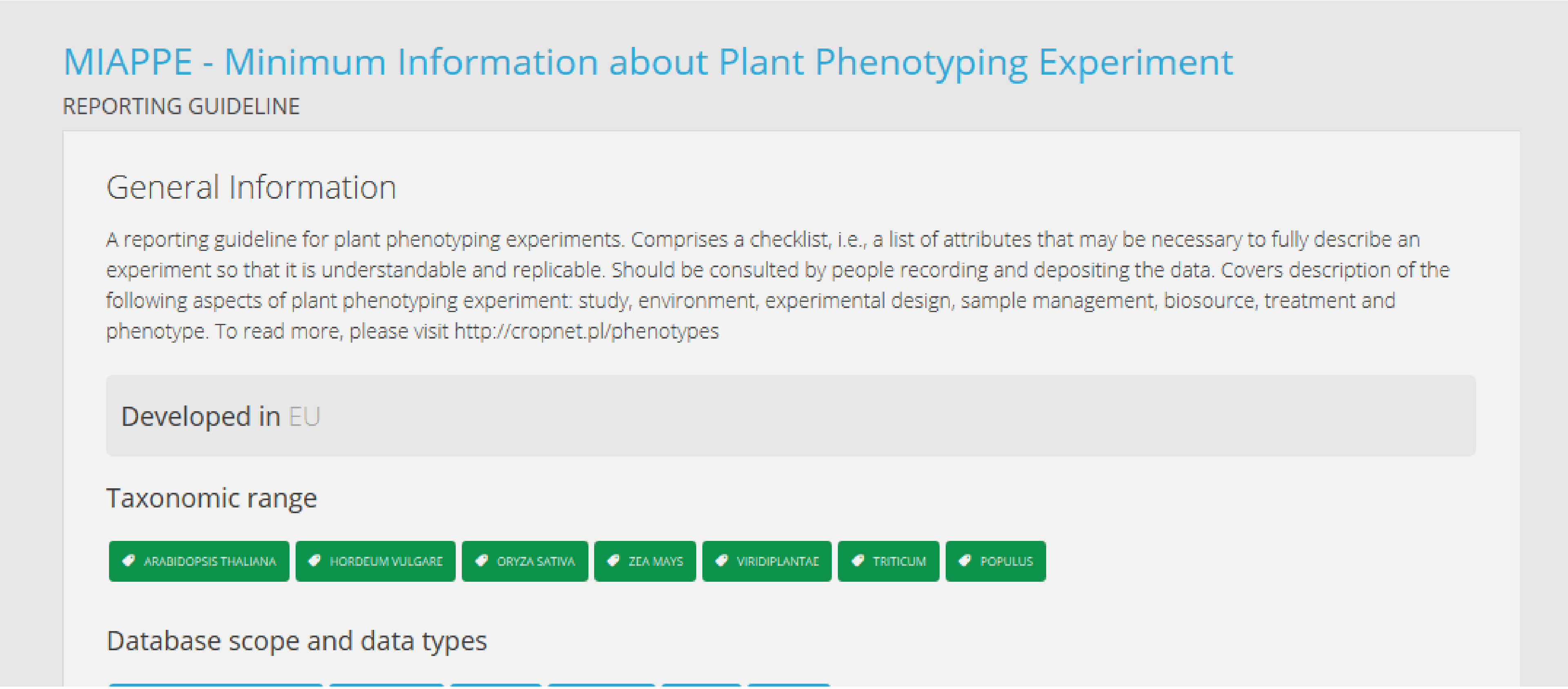
Partly compiled with the support of

NAR Database Issue and *DATABASE* journals

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Minimum Information about Plant Phenotypic Experiment (MIAPPE)

http://www.biosharing.org/bsg-000543



MIAPPE - Minimum Information about Plant Phenotyping Experiment

REPORTING GUIDELINE

General Information

A reporting guideline for plant phenotyping experiments. Comprises a checklist, i.e., a list of attributes that may be necessary to fully describe an experiment so that it is understandable and replicable. Should be consulted by people recording and depositing the data. Covers description of the following aspects of plant phenotyping experiment: study, environment, experimental design, sample management, biosource, treatment and phenotype. To read more, please visit http://cropnet.pl/phenotypes

Developed in EU

Taxonomic range

ARABIDOPSIS THALIANA · HORDEUM VULGARE · ORYZA SATIVA · ZEA MAYS · VIRIDIPLANTAE · TRITICUM · POPULUS

Database scope and data types

Format wymiany danych

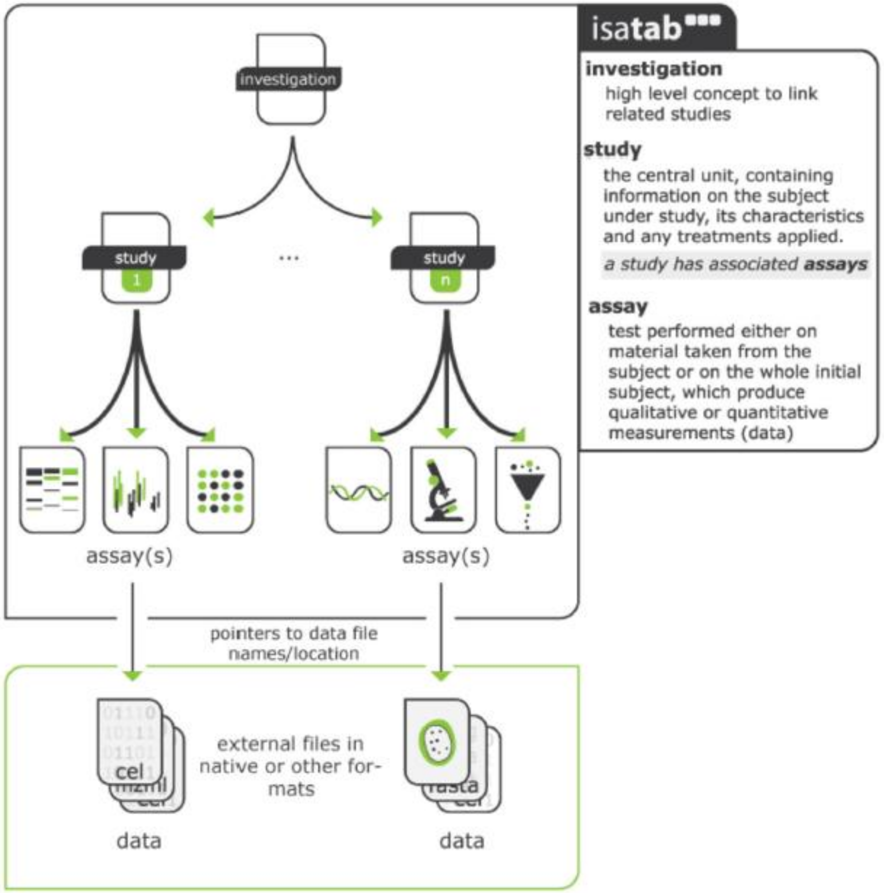
ISA-TAB: Investigation/Study/Assay (http://www.isa-tools.org/)

Investigation

Study

Assay

Data files



investigation
high level concept to link related studies

study
the central unit, containing information on the subject under study, its characteristics and any treatments applied.
*a study has associated **assays***

assay
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

Sansone & Rocca-Serra (2012)

Pliki tekstowe

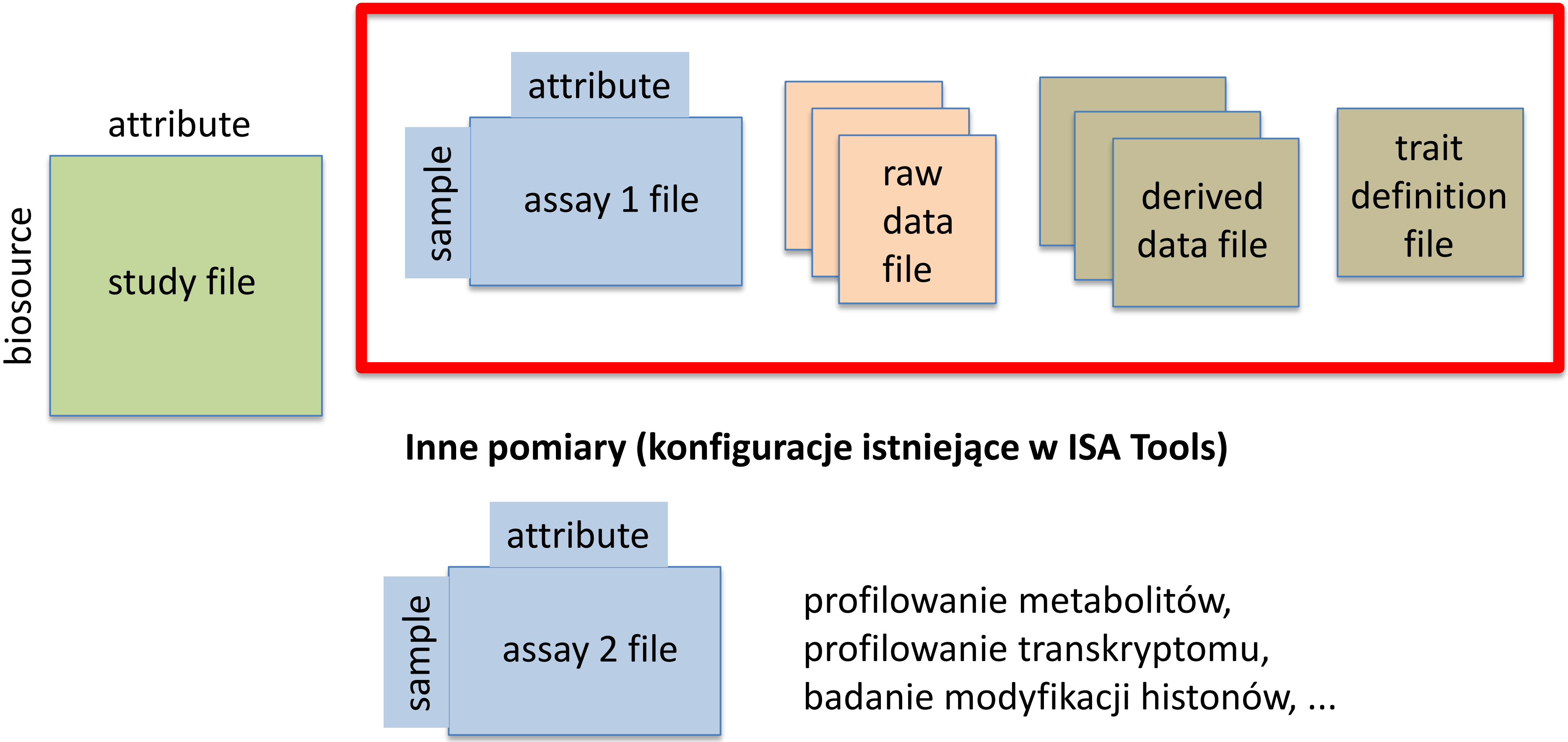
Struktury:
- source/sample
- factor
- characteristics
- protocol
- parameter
- unit
- extract
- comment

Wsparcie dla adnotacji

Narzędzia:
ISA Software Suite

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Format wymiany danych



Pomiary fenotypowe (konfiguracja fenotypowa ISA-TAB)

attribute

biosource

study file

attribute

sample

assay 1 file

raw data file

derived data file

trait definition file

Inne pomiary (konfiguracje istniejące w ISA Tools)

attribute

sample

assay 2 file

profilowanie metabolitów,
profilowanie transkryptomu,
badanie modyfikacji histonów, ...

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Zastosowania

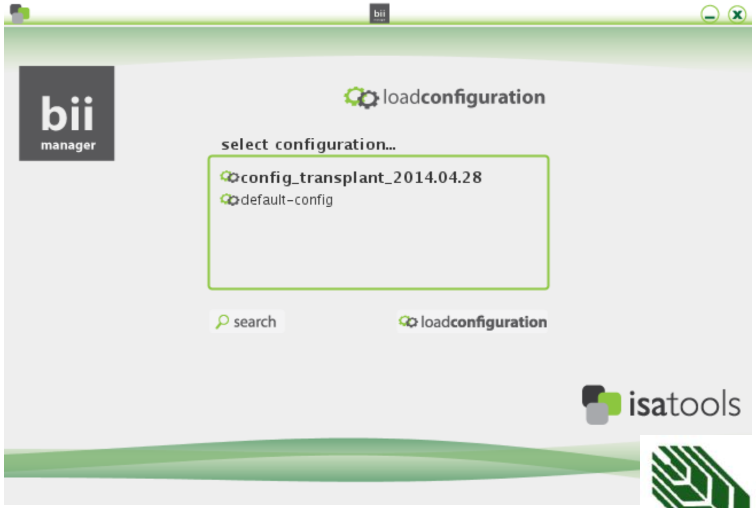
Baza danych oparta na oprogramowaniu [BII Data Manager]

[http://www.igr.poznan.pl/bb/bii]

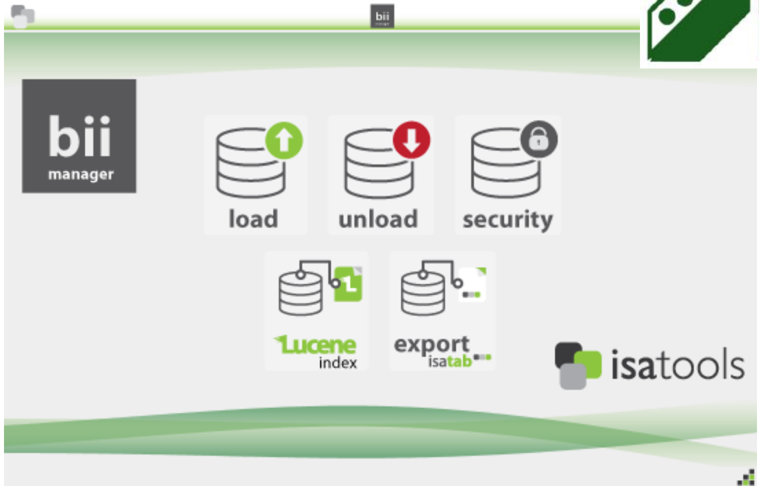
Dane w formacie ISA-TAB

Filtrowanie

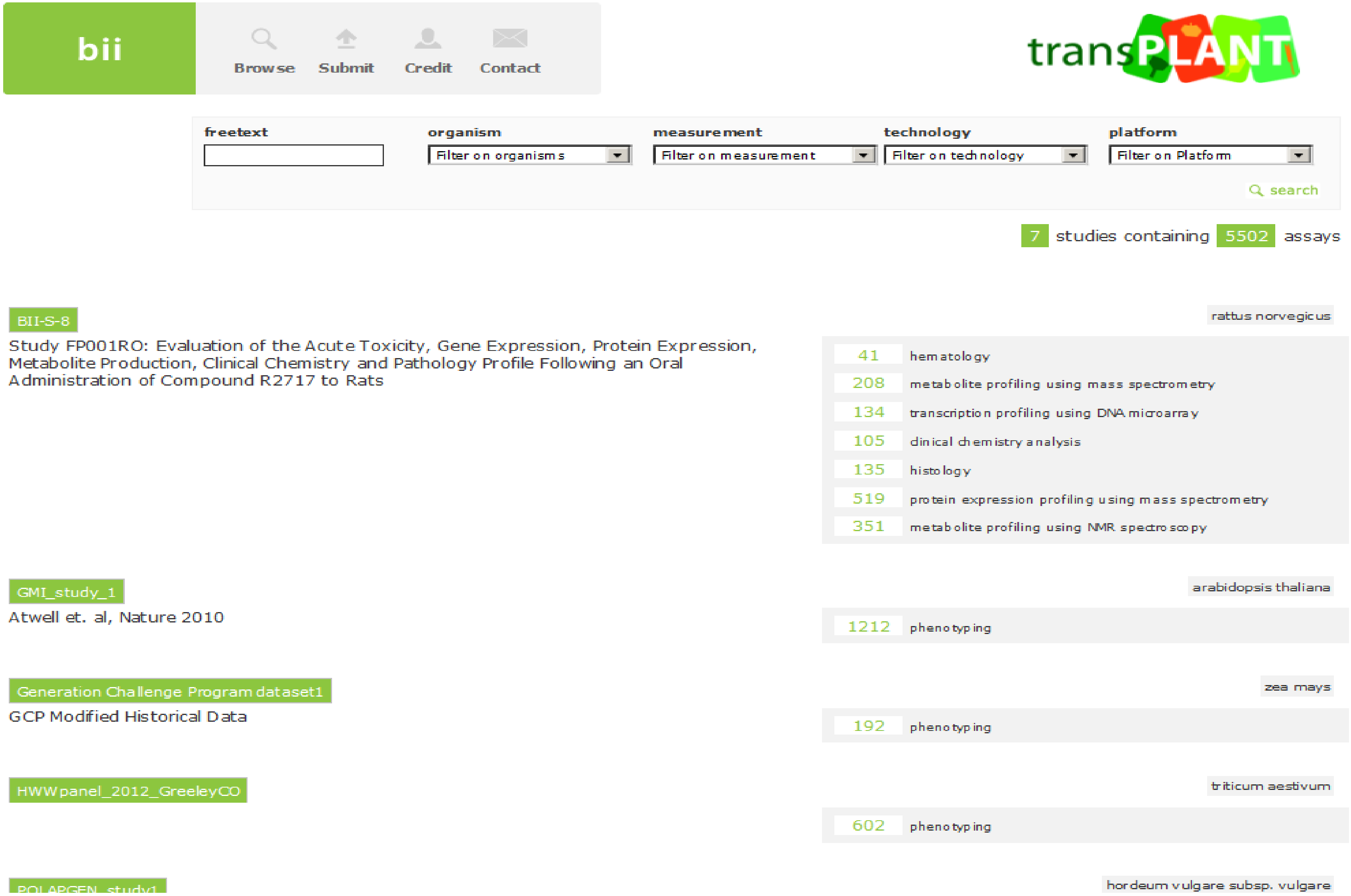
Wyszukiwanie metadanych



config_transplant_2014.04.28
default-config



The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area "Infrastructures". Contract number 283496.



7 studies containing 5502 assays

BII-S-8

Study FP001RO: Evaluation of the Acute Toxicity, Gene Expression, Protein Expression, Metabolite Production, Clinical Chemistry and Pathology Profile Following an Oral Administration of Compound R2717 to Rats

rattus norvegicus

41 hematology
208 metabolite profiling using mass spectrometry
134 transcription profiling using DNA microarray
105 clinical chemistry analysis
135 histology
519 protein expression profiling using mass spectrometry
351 metabolite profiling using NMR spectroscopy

GMI_study_1

Atwell et. al, Nature 2010

arabidopsis thaliana

1212 phenotyping

Generation Challenge Program dataset1

GCP Modified Historical Data

zea mays

192 phenotyping

HWWpanel_2012_GreeleyCO

triticum aestivum

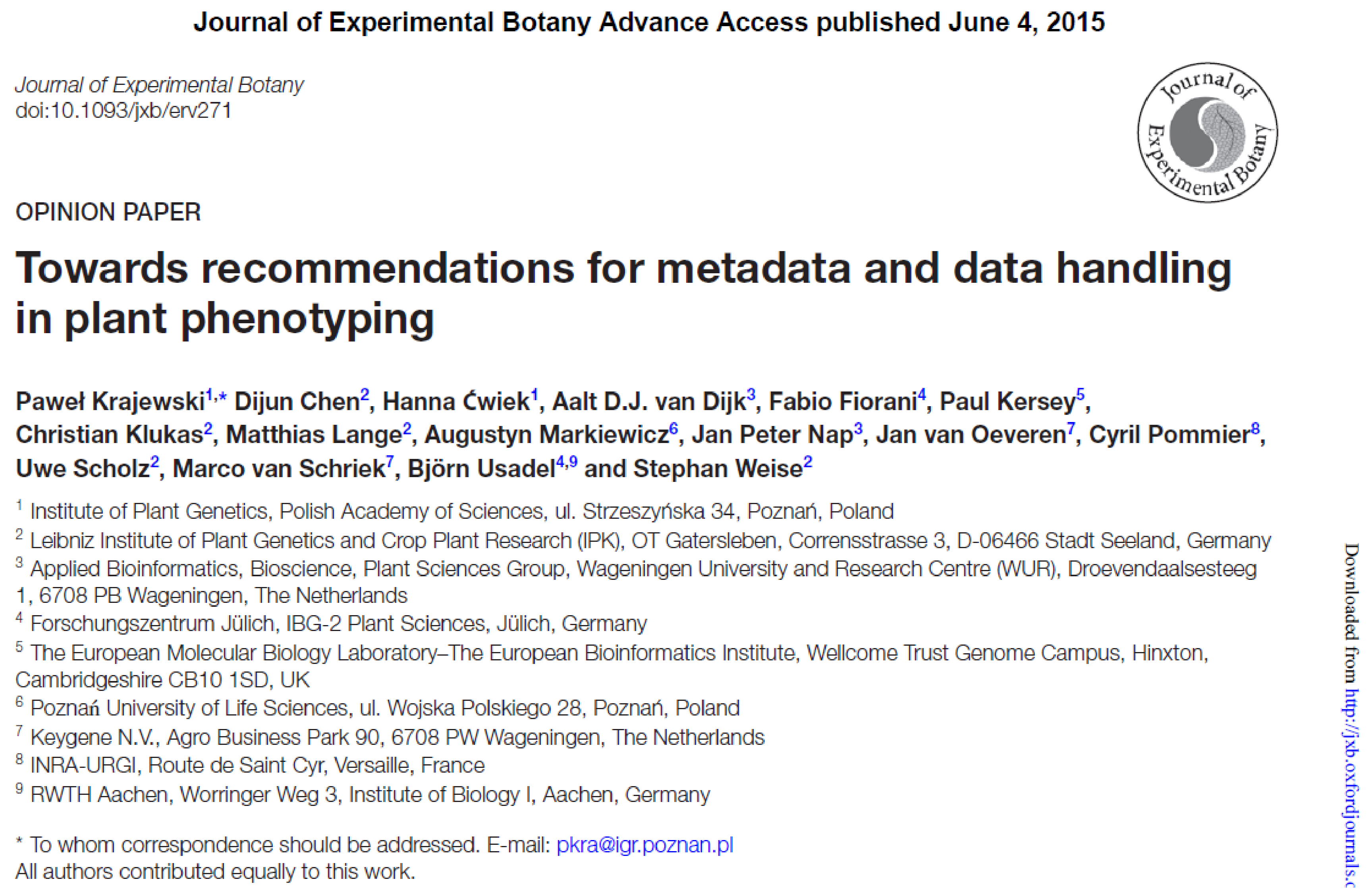
602 phenotyping

POLAPGEN_study1

hordeum vulgare subsp. vulgare

www.transplantdb.eu

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.



Journal of Experimental Botany Advance Access published June 4, 2015

Journal of Experimental Botany
doi:10.1093/jxb/erv271

OPINION PAPER

Towards recommendations for metadata and data handling in plant phenotyping

Paweł Krajewski[1,*] Dijun Chen[2], Hanna Ćwiek[1], Aalt D.J. van Dijk[3], Fabio Fiorani[4], Paul Kersey[5], Christian Klukas[2], Matthias Lange[2], Augustyn Markiewicz[6], Jan Peter Nap[3], Jan van Oeveren[7], Cyril Pommier[8], Uwe Scholz[2], Marco van Schriek[7], Björn Usadel[4,9] and Stephan Weise[2]

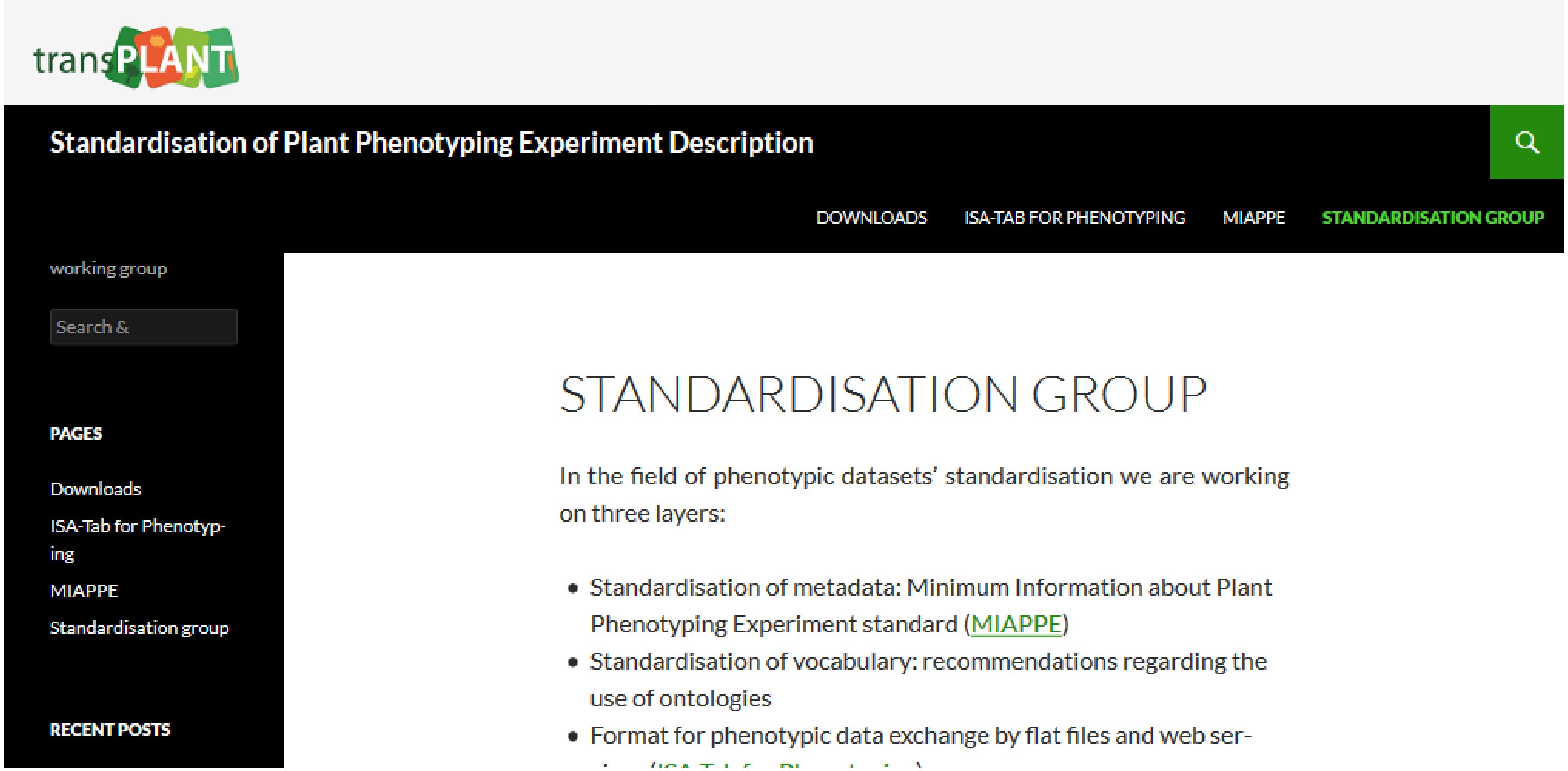
[1] Institute of Plant Genetics, Polish Academy of Sciences, ul. Strzeszyńska 34, Poznań, Poland
[2] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), OT Gatersleben, Corrensstrasse 3, D-06466 Stadt Seeland, Germany
[3] Applied Bioinformatics, Bioscience, Plant Sciences Group, Wageningen University and Research Centre (WUR), Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[4] Forschungszentrum Jülich, IBG-2 Plant Sciences, Jülich, Germany
[5] The European Molecular Biology Laboratory–The European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK
[6] Poznań University of Life Sciences, ul. Wojska Polskiego 28, Poznań, Poland
[7] Keygene N.V., Agro Business Park 90, 6708 PW Wageningen, The Netherlands
[8] INRA-URGI, Route de Saint Cyr, Versaille, France
[9] RWTH Aachen, Worringer Weg 3, Institute of Biology I, Aachen, Germany

* To whom correspondence should be addressed. E-mail: pkra@igr.poznan.pl
All authors contributed equally to this work.

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.

Informacje

cropnet.pl/phenotypes



transPLANT

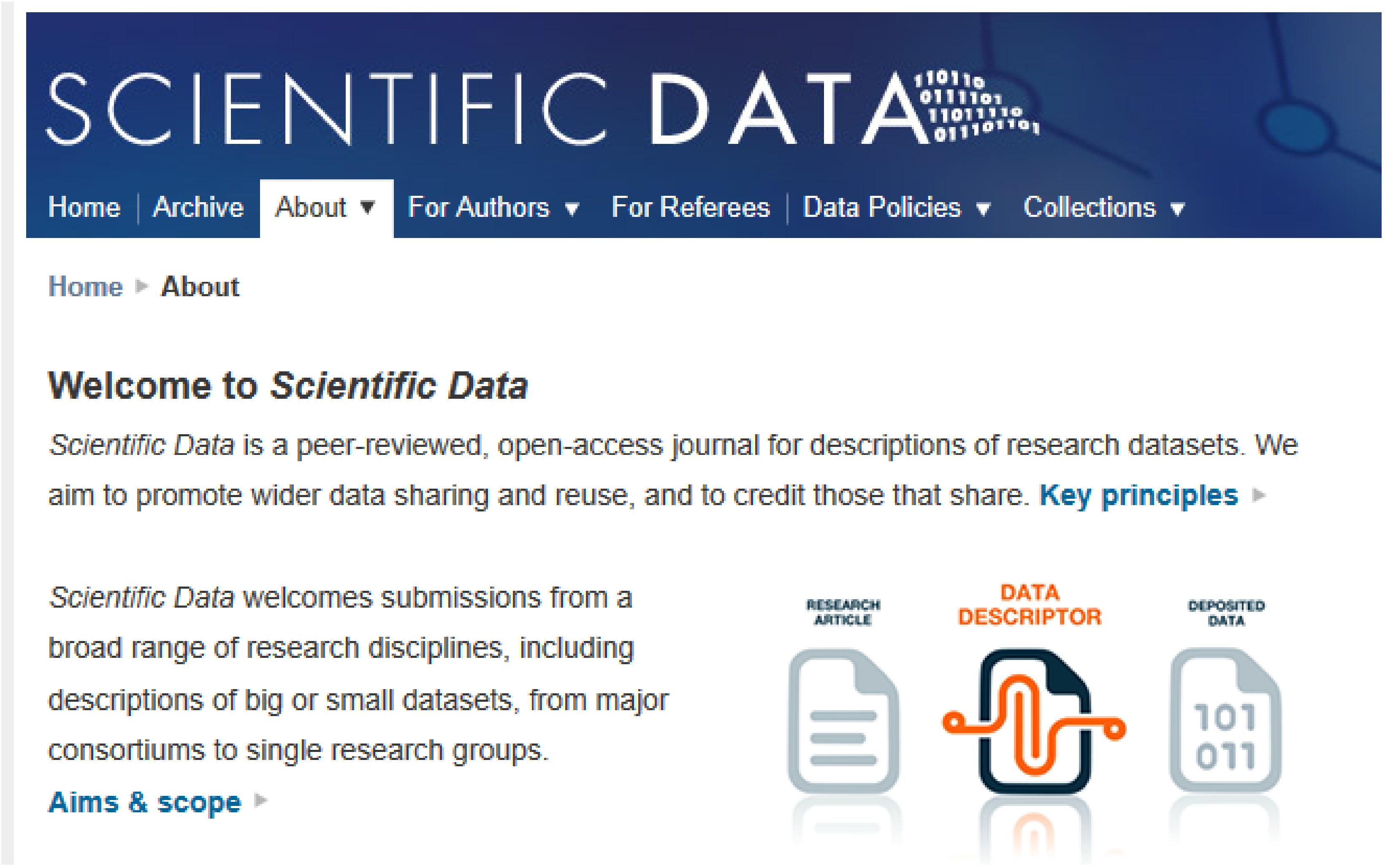
Standardisation of Plant Phenotyping Experiment Description

STANDARDISATION GROUP

In the field of phenotypic datasets' standardisation we are working on three layers:

- Standardisation of metadata: Minimum Information about Plant Phenotyping Experiment standard (MIAPPE)
- Standardisation of vocabulary: recommendations regarding the use of ontologies
- Format for phenotypic data exchange by flat files and web ser-

Publikacja danych



Welcome to *Scientific Data*

Scientific Data is a peer-reviewed, open-access journal for descriptions of research datasets. We aim to promote wider data sharing and reuse, and to credit those that share. **Key principles** ▸

Scientific Data welcomes submissions from a broad range of research disciplines, including descriptions of big or small datasets, from major consortiums to single research groups.

Aims & scope ▸

IGR PAN

Hanna Ćwiek
Wojtek Frohmberg
Augustyn Markiewicz

PSNC

Cezary Mazurek
Raul Palma
Fryderyk Raczyk

EMBL-EBI

Paul Kersey
(Ensembl, Elixir)

KeyGene

Jan van Oeveren
Marco van Shriek

JÜLICH
FORSCHUNGSZENTRUM

Björn Usadel
Fabio Fiorani
Hendrik Poorter
(EPPN, DPPN)

INRA

Hadi Quesneville
Cyril Pommier
(RDA Wheat Data WG)

IPK
GATERSLEBEN

Matthias Lange
Astrid Junker
Uwe Scholz
Daniel Arendt
Stephan Weise
Christian Klukas
Dijun Chen

GMI
GREGOR MENDEL INSTITUTE
OF MOLECULAR PLANT BIOLOGY

Magnus Nordborg
Ümit Seren

PLANT RESEARCH
INTERNATIONAL
WAGENINGEN UR

Aalt-Jan van Dijk
Jan Peter Nap

Oxford
e-Research
Centre

isatools

Susanna Sansone
Philippe Rocca-Serra

transPLANT

www.transplantdb.eu

The transPLANT project is funded by the European Commission within its 7th Framework Programme under the thematic area “Infrastructures”. Contract number 283496.