



arnes 



**NI4OS**  
National Initiatives for Open Science in Europe

# Primer ravnanja z biološkimi in biotehnološkimi podatki v okviru raziskovalnih projektov na NIB

dr. Marko Petek, NIB

Mreža znanja 2021, od 23. do 30. novembra



# Raziskovalni podatki na Oddelku za biotehnologijo in sistemsko biologijo

- Surovi podatki
- Metapodatki
- Vmesni in končni rezultati obdelave podatkov in modeliranja
- Publikacije (docx, pptx, pdf...)

# Načrt ravnanja s projektnimi podatki (“data management plan”)

	Synthetic biology constructs / Phytobricks	Transcriptomics	Metabolomics	Electroantennography	Plant phenotyping	Promoter transcriptional activity
<b>Generated by</b>	CSIC, EI, TUDA	NIB, EI	TUDA, CSIC	TUDA	CSIC, EI	CSIC
<b>Format of generated data</b>	GenBank	fastq, hdf5	DATA.MS, NMReDATA, Bruker data format, Shimadzu data format (qgd)	txt(ASCII), jpeg, eag	jpeg, txt(ASCII)	txt(ASCII), jpeg
<b>Raw data storage at</b>	CSIC, EI, TUDA, GeneBank	NIB, EI, SRA/ArrayExpress/ GEO	TUDA, CSIC, MetaboLights, MolCheck	TUDA, FAIRDOMHub	CSIC, EI, FAIRDOMHub	CSIC, FAIRDOMHub
<b>Expected data size</b>	< 1 GB	< 1800 GB	< 30 GB	< 5 GB	< 10 GB	< 1 GB
<b>Analysed by</b>	CSIC, EI, TUDA	NIB, EI	TUDA, CSIC	TUDA	CSIC, EI	CSIC
<b>Analysed data storage at</b>	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	GB Elements Database, FAIRDOMHub
<b>Minimal information requirements</b>	MIRIAM	MIAME	CIMR	MINI	MIAPPE	/
<b>Standards, formats</b>	Genebank, SBOL data and visual	MAGE-ML	mzML, mzQuantML, nmrML	NWB	/	/
<b>Ontologies and vocabularies used</b>	SBOL	GO, KEGG, InterPro, MapMan	CHEBI, KEGG, MapMan	OEN	PO, TO, CO	SBOL
<b>SOPs stored at</b>	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub
<b>Scripts stored at</b>	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub	FAIRDOMHub

# Podatkovni repozitoriji

## SPECIALIZIRANI REPOZITORIJI ZA SUROVE BIOLOŠKE PODATKE

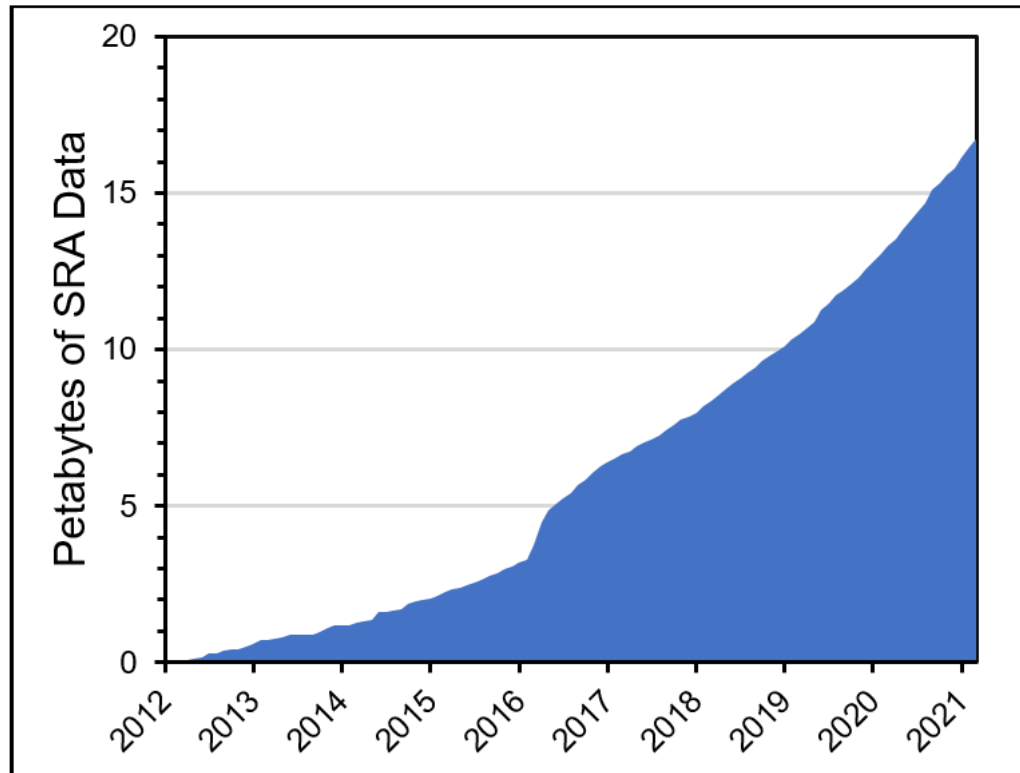


## SPLOŠNI REPOZITORIJI ZA PODATKE

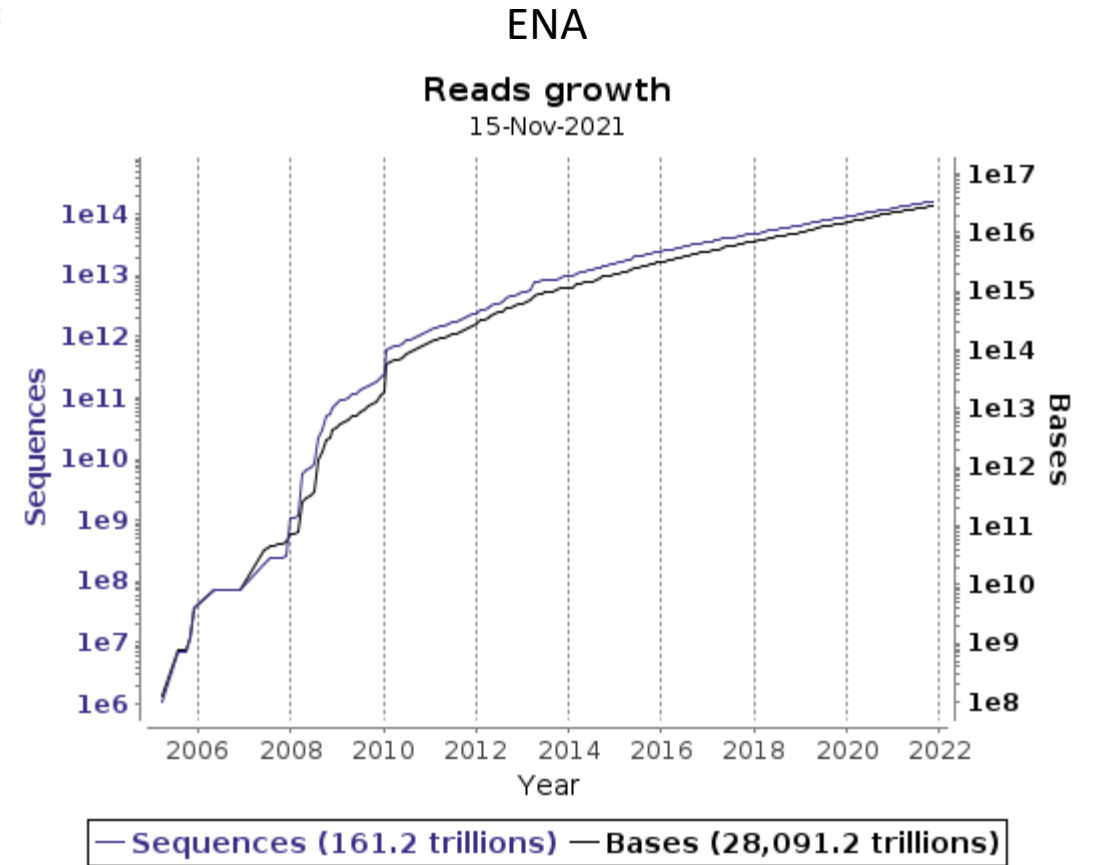


# Podatkovni repozitoriji

Rast količine podatkov visokozmogljivostnega sekvenciranja DNA (in RNA):



Vir: NCBI Insights  
(<https://ncbiinsights.ncbi.nlm.nih.gov/2021/08/09/espsss-workshop/#more-6180>)



Vir: EBI ENA  
(<https://www.ebi.ac.uk/ena/browser/about/statistics>)

# Deponiranje surovih podatkov v javne repozitorije

NCBI Resources How To markop

BioProject

Advanced Browse by Project attributes

## Leptinotarsa decemlineata midgut RNASeq

The goal of the project was to investigate gene expression in Colorado potato beetle larval midguts reared (healthy) potato plants.

Accession	PRJNA400633
Data Type	Raw sequence reads
Scope	Multispecies
Publications	1. Published online: Schoville S <i>et al.</i> , "A model species for agricultural pest genomics: the genome of <i>Leptinotarsa decemlineata</i> (Coleoptera: Chrysomelidae)", <i>Scientific Reports</i> , 2018;8(1)
Grants	"Biotechnology and Plant Systems Biology" (Grant ID P4-0165, Slovenian Research Agency)
Submission	Registration date: 29-Aug-2017 National Institute of Biology
Related Resources	<ul style="list-style-type: none"> <li>Growth and defense trade-offs in multitrophic interaction between potato and its two major herbivores</li> <li>Using RNAi and SysBio approaches for validation of insecticide targets in CPB guts</li> <li>Colorado Potato Beetle Genome Project</li> </ul>
Relevance	Agricultural

### Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
SRA Experiments	3
OTHER DATASETS	
BioSample	3

SRA Data Details	
Parameter	Value
Data volume, Gbases	9
Data volume, Mbytes	7866

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA400633>

NCBI Resources How To markop

SRA

Advanced

Send to:

Related information

Full

Send to:

Related information

Recent activity

NCBI Site map All databases Search

**Sequence Read Archive**

Main Browse Search Download Submit Software Trace Archive Trace BLAST

Studies Samples Analyses Run Browser Run Selector Provisional SRA

RNA-seq of *Leptinotarsa decemlineata*: larval midgut HEALTHY-reared (SRR5989325)

Metadata Analysis Reads Data access

Run	Spots	Bases	Size	Published	Access Type
SRR5989325	95.9M	4.8Gbp	4.1G	2017-09-24	public

Quality graph (bigger)

This run has 1 read per spot:

L=50, 100%

Legend

Experiment	Library Name	Platform	Strategy	Source	Selection	Layout
SRX3145066	21_HEALTHY	ABI Solid	RNA-Seq	TRANSCRIPTOMIC	RANDOM	SINGLE

Show design

Biosample	Sample Description	Organism	Links
SAMN07572226 (SRS2477339)		<a href="#">Leptinotarsa decemlineata</a>	<a href="#">PRJNA400633</a>

Bioproject	SRA Study	Title
<a href="#">PRJNA400633</a>	<a href="#">SRP116557</a>	Leptinotarsa decemlineata midgut RNASeq

Abstract:

The goal of the project was to investigate gene expression in Colorado potato beetle larval midguts reared on virus infected vs non-infected (healthy) potato plants.

Submitted by: National Institute of Biology

**Study:** *Leptinotarsa decemlineata* midgut RNASeq  
[PRJNA400633](#) • [SRP116557](#) • [All experiments](#) • [All runs](#)

[hide Abstract](#)

The goal of the project was to investigate gene expression in Colorado potato beetle larval midguts reared on virus infected vs non-infected (healthy) potato plants.

**Sample:** Midguts of *Leptinotarsa decemlineata* larvae reared continuously on potted non-infected potato plants.  
[SAMN07572226](#) • [SRS2477339](#) • [All experiments](#) • [All runs](#)  
 Organism: [Leptinotarsa decemlineata](#)

**Library:**  
 Name: 21\_HEALTHY  
 Instrument: AB SOLiD 4 System  
 Strategy: RNA-Seq  
 Source: TRANSCRIPTOMIC  
 Selection: RANDOM  
 Layout: SINGLE

**Spot descriptor:**  
 1 forward

**Links:**

Runs: 1 run, 95.9M spots, 4.8G bases, 3.8Gb

Run	# of Spots	# of Bases	Size	Published
<a href="#">SRR5989325</a>	95,930,453	4.8G	3.8Gb	2017-09-24

ID: 4441056



**Kako na NIB-u ravnamo s podatki preden jih deponiramo v javnih repozitorijih?**



# Kako na NIB-u ravnamo s podatki preden jih deponiramo v javnih repozitorijih?

- Surovi podatki
- Metapodatki



mrežni strežnik z  
varnostno kopijo in  
omejenim dostopom

- Vmesni in končni rezultati obdelave podatkov in modeliranja
- Publikacije (docx, pptx, pdf...)

# Kako na NIB-u ravnamo s podatki preden jih deponiramo v javnih repozitorijih?

- Surovi podatki
- Metapodatki
- Vmesni in končni rezultati obdelave podatkov in modeliranja
- Publikacije (docx, pptx, pdf...)



mrežni strežnik z varnostno kopijo in omejenim dostopom



BUFO



IBIS2



Hansel Gretel

linux strežniki za obdelavo podatkov



# Kako na NIB-u ravnamo s podatki preden jih deponiramo v javnih repozitorijih?

- Surovi podatki
- Metapodatki
- Vmesni in končni rezultati obdelave podatkov in modeliranja
- Publikacije (docx, pptx, pdf...)



mrežni strežnik z varnostno kopijo in omejenim dostopom

drugi mrežni strežnik z varnostno kopijo



bufo



IBIS2



Hansel Gretel

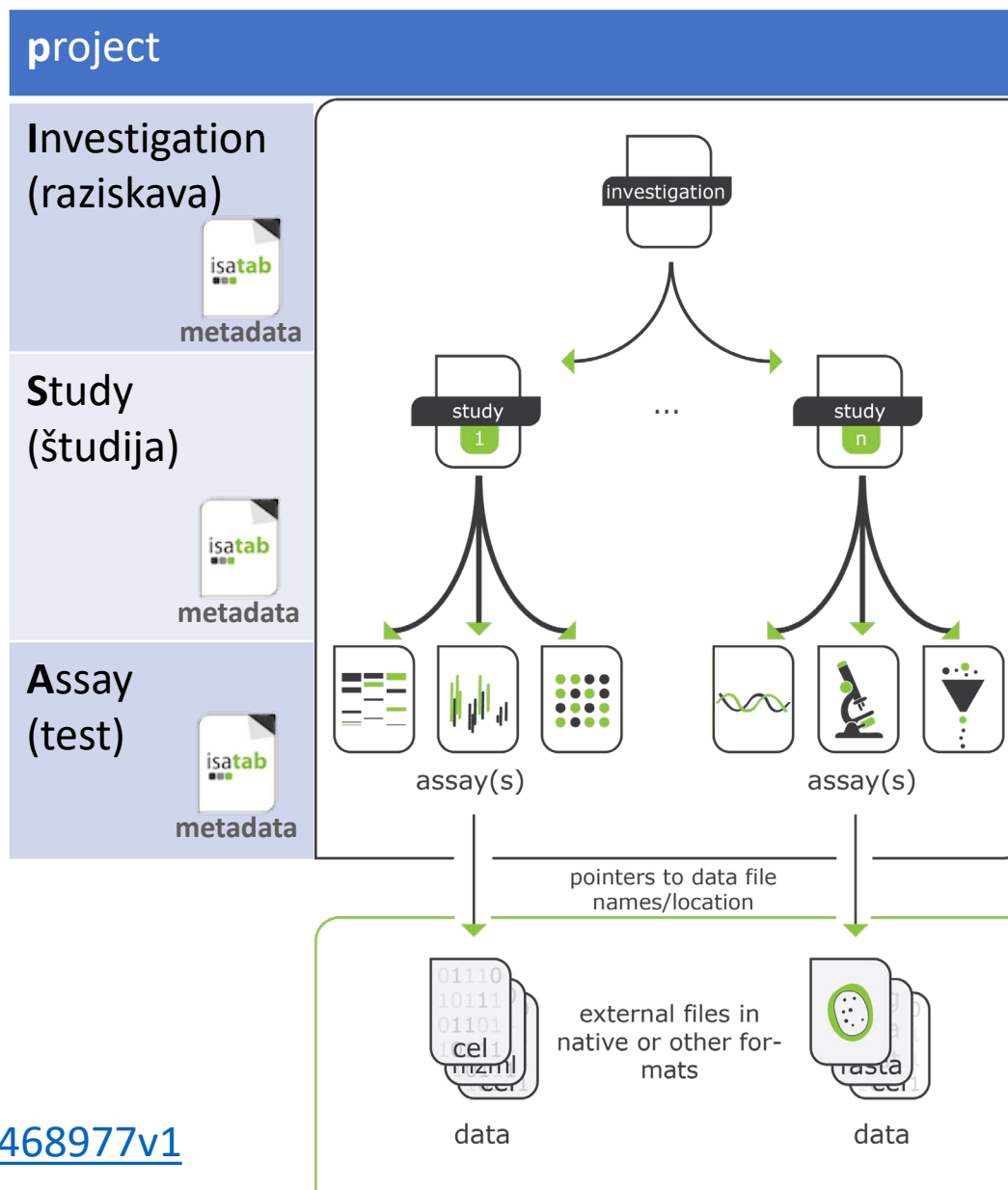
linux strežniki za obdelavo podatkov

# pISA-tree

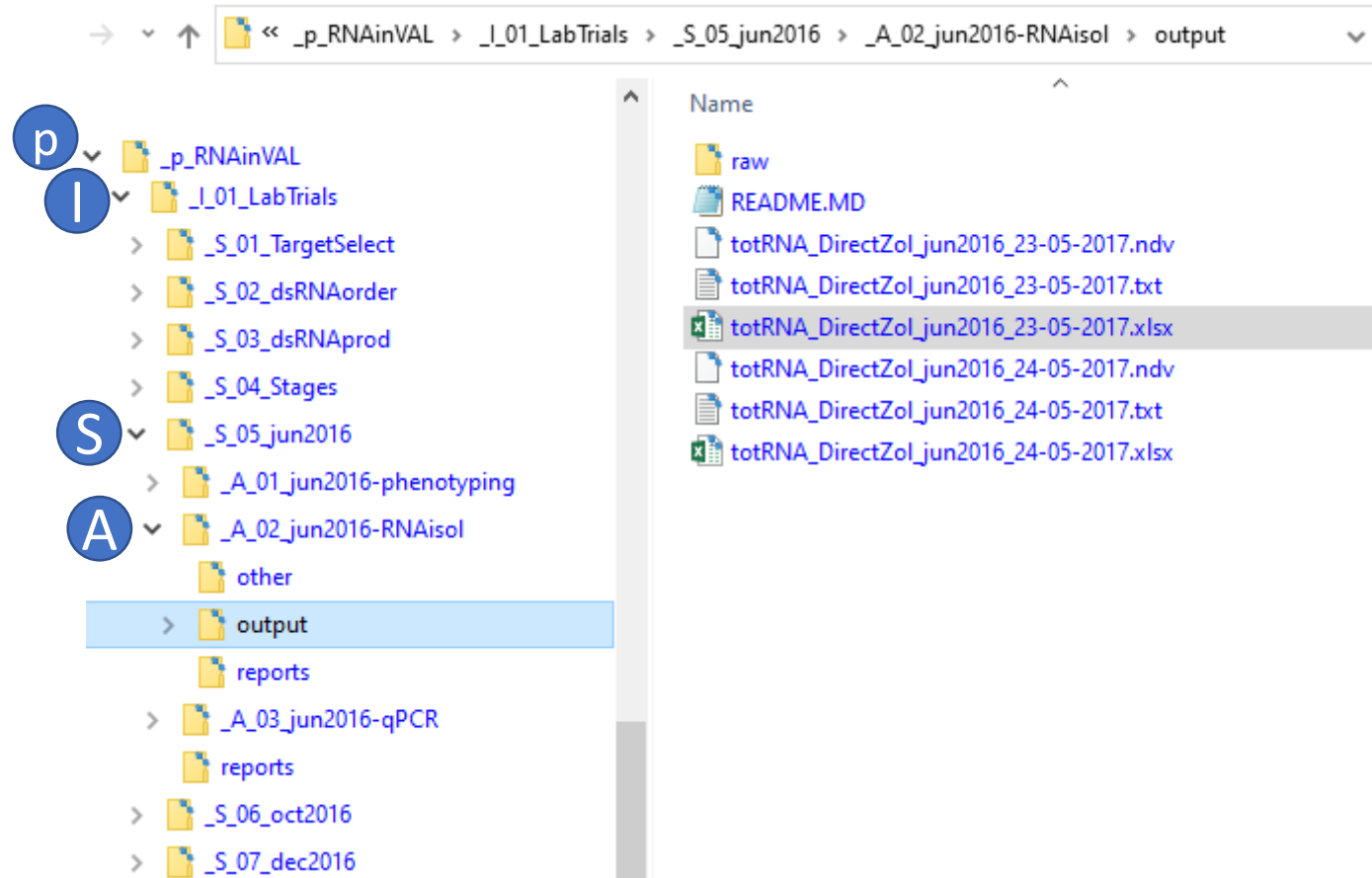
- Sistem za organizacijo projektnih podatkov (\*.bat skripte)
- Uporabljamo datotečne mape urejene v drevesno strukturo po **ISA standardu**
- metapodatkovne txt datoteke v **ISA-Tab formatu**

<https://github.com/NIB-SI/pISA-tree>

<https://www.biorxiv.org/content/10.1101/2021.11.18.468977v1>



# pISA-tree: primer lokale strukture



# pISA-tree: primer lokalne strukture

The screenshot shows a Windows File Explorer window with the following path: `<< _p_RNAinVAL > _I_01_LabTrials > _S_05_jun2016 > _A_02_jun2016-RNAisol`. The left sidebar shows a tree view of folders, with `_A_02_jun2016-RNAisol` selected. The main pane displays a table of files and folders:

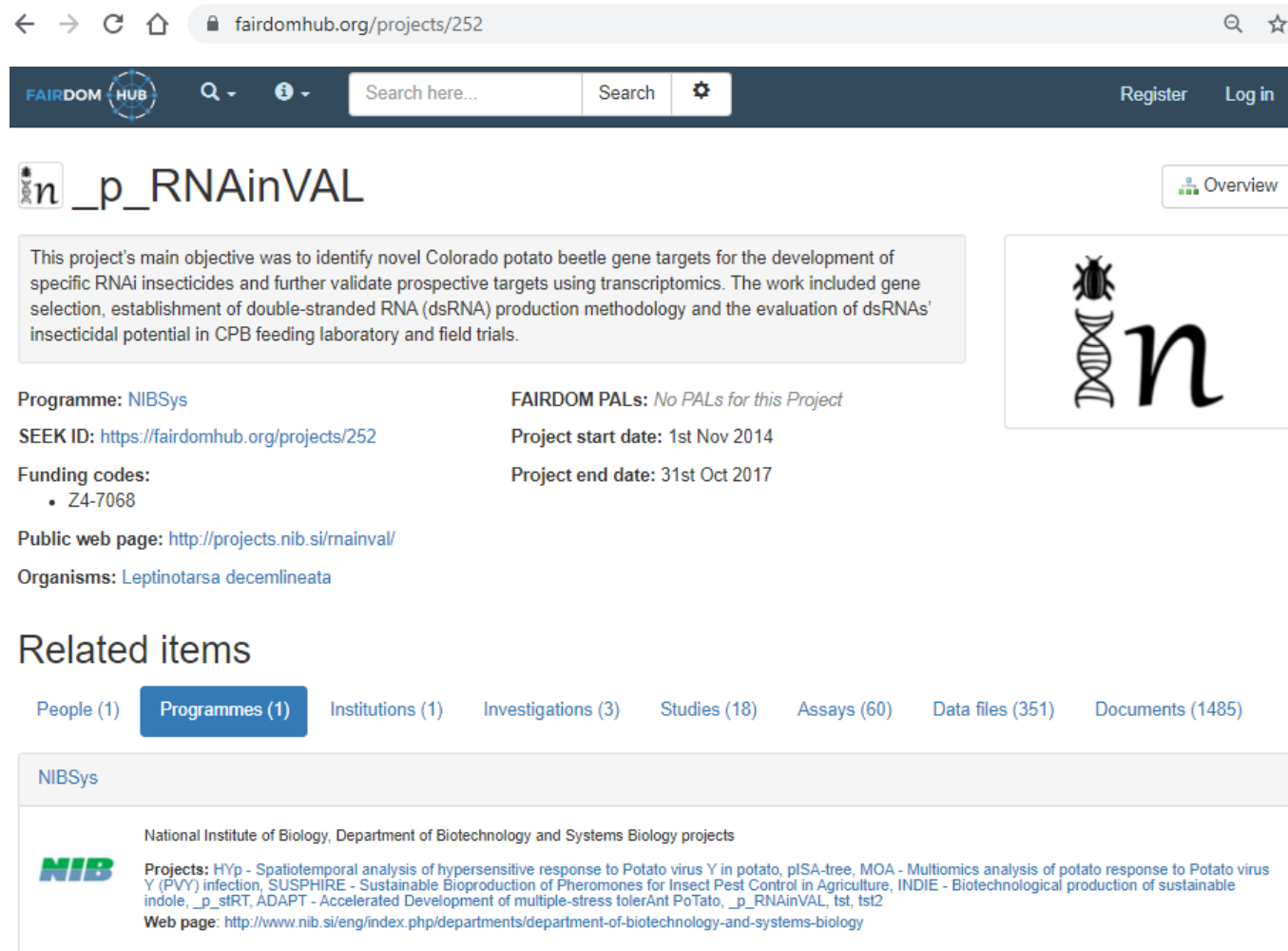
Name	Date modified	Type	Size
other	13. 05. 2021 17:22	File folder	
output	13. 05. 2021 17:16	File folder	
reports	10. 05. 2021 16:27	File folder	
<b>_ASSAY_METADATA.TXT</b>	12. 11. 2021 15:33	Text Document	1 KB
analytes.xlsx	17. 05. 2021 10:34	Microsoft Excel Work...	16 KB
README.MD	10. 05. 2021 16:27	MD File	1 KB
showMetadata.bat	1. 04. 2020 14:26	Windows Batch File	1 KB
showTree.bat	1. 04. 2020 14:26	Windows Batch File	1 KB

A Notepad window titled `_ASSAY_METADATA.TXT - Notepad` is open, displaying the following metadata:

```
File Edit Format View Help
Assay: _A_02_jun2016-RNAisol
Short Name: 02_jun2016-RNAisol
Assay Class: WET
Assay Type: RNAisol
Title: RNA isolation and pre-qPCR sample prep
Description: RNA isolation and sample preparation for qPCR
pISA Assay creation date: 2021-05-10
pISA Assay creator: Marko Petek
Lab manager: Marko Petek
Phenodata: ../../phenodata_20210113.txt
Featuredata:
RNA ID: $_RNA
Homogenisation protocol: handheld electric rotor-stator homogenization in tubes with Trizol
Date Homogenisation: 2017-04-26
Isolation Protocol: ZymoRNA
Date Isolation: 2017-04-26
Storage RNA: CU470
Dnase treatment protocol: Zymo DNase I
Dnase ID: $_DNase
Date DNase_treatment: 2017-05-10
Storage_DNase_treated: NA
Operator: Marko Petek, Katja Stare
cDNA ID: $_cDNA
DateRT: 2017-05-10
Notes:
Data:
```

# pISA-tree --> FAIRDOMHub.org

- FAIRDOMHub
  - uporablja ISA format
  - na voljo večina licenc za odprto kodo in podatke (Creative Commons, Open Data Commons, ...)
  - omogoča pridobitev DOI za podatke
- Prenos v programskem okolju R s paketom *seekr* (<https://github.com/NIB-SI/seekr>)



The screenshot shows the FAIRDOMHub.org website interface. The browser address bar displays 'fairdomhub.org/projects/252'. The page title is 'indole\_p\_RNAinVAL'. A description box states: 'This project's main objective was to identify novel Colorado potato beetle gene targets for the development of specific RNAi insecticides and further validate prospective targets using transcriptomics. The work included gene selection, establishment of double-stranded RNA (dsRNA) production methodology and the evaluation of dsRNAs' insecticidal potential in CPB feeding laboratory and field trials.' Metadata includes: Programme: NIBSys; FAIRDOM PALs: No PALs for this Project; SEEK ID: https://fairdomhub.org/projects/252; Project start date: 1st Nov 2014; Funding codes: Z4-7068; Project end date: 31st Oct 2017; Public web page: http://projects.nib.si/rnainval; Organisms: Leptinotarsa decemlineata. A 'Related items' section shows filters for People (1), Programmes (1), Institutions (1), Investigations (3), Studies (18), Assays (60), Data files (351), and Documents (1485). The 'NIBSys' section identifies the National Institute of Biology, Department of Biotechnology and Systems Biology projects, listing 'indole\_p\_RNAinVAL' as a project.





# Zahvala



prof. dr. Kristina Gruden



doc. dr. Špela Baebler



dr. Tjaša Lukan



prof. dr. Andrej Blejec



dr. Živa Ramšak



Katja Stare



dr. Maja Zagorščak



dr. Anna Coll Rius



Valentina Levak

moj email: [marko.petek@nib.si](mailto:marko.petek@nib.si)

Hvala za pozornost!