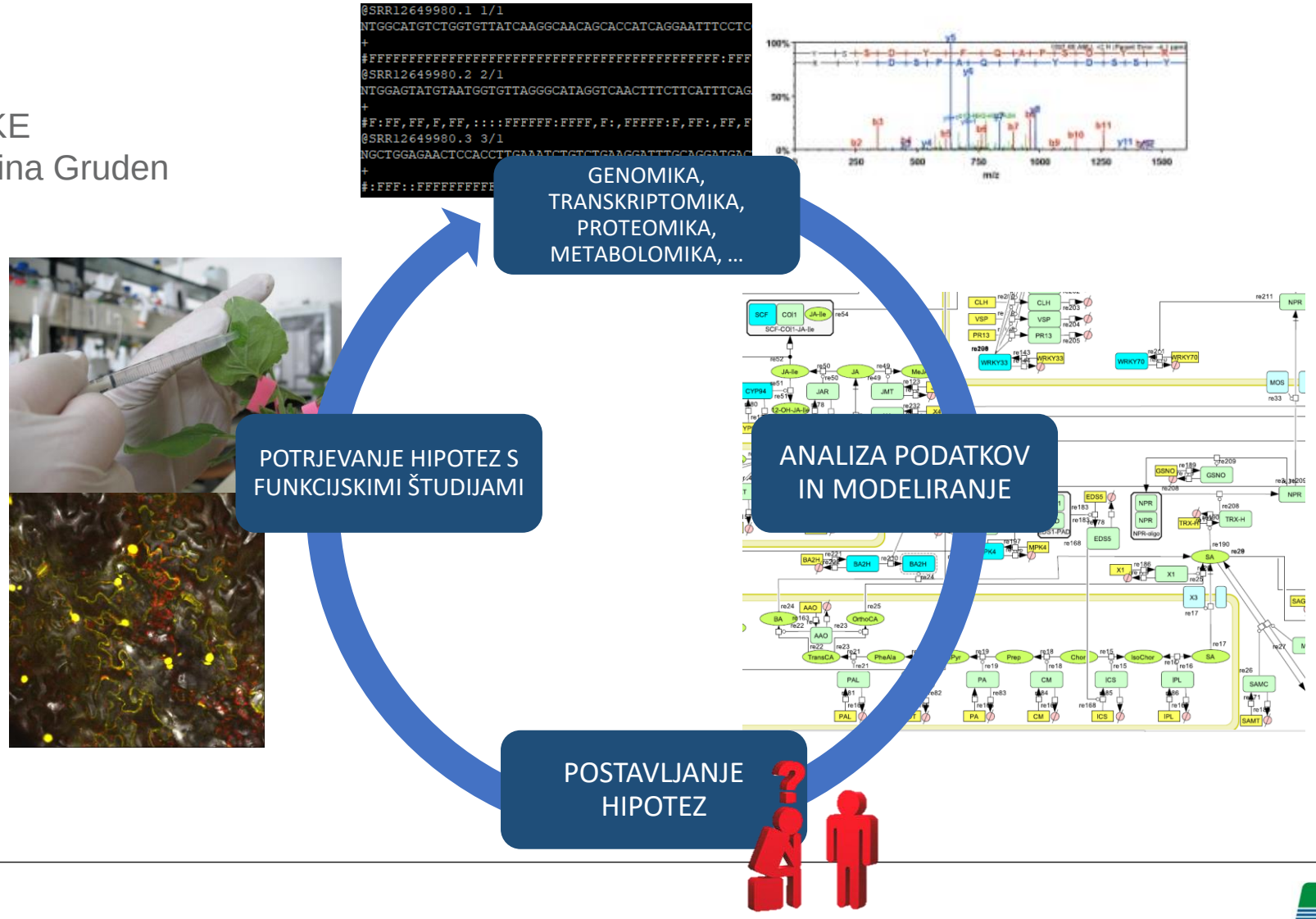
Primer ravnanja z biološkimi in biotehnološkimi podatki v okviru raziskovalnih projektov na NIB

dr. Marko Petek, NIB

Mreža znanja 2021, od 23. do 30. novembra

Raziskave na Oddelku za biotehnologijo in sistemsko biologijo

Delovni sklop: OMIKE
Vodja: prof. dr. Kristina Gruden



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

Podatkovni repozitoriji

SPECIALIZIRANI REPOZITORIJI ZA SUROVE BIOLOŠKE PODATKE

GEO Gene Expression Omnibus

ENA European Nucleotide Archive

SRA

MetaboLights

PubChem

e!DAL Plant Genomics & Phenomics Research Data Repository

PRIDE

EMPIAR Electron Microscopy Public Image Archive

SPLOŠNI REPOZITORIJI ZA PODATKE

zenodo

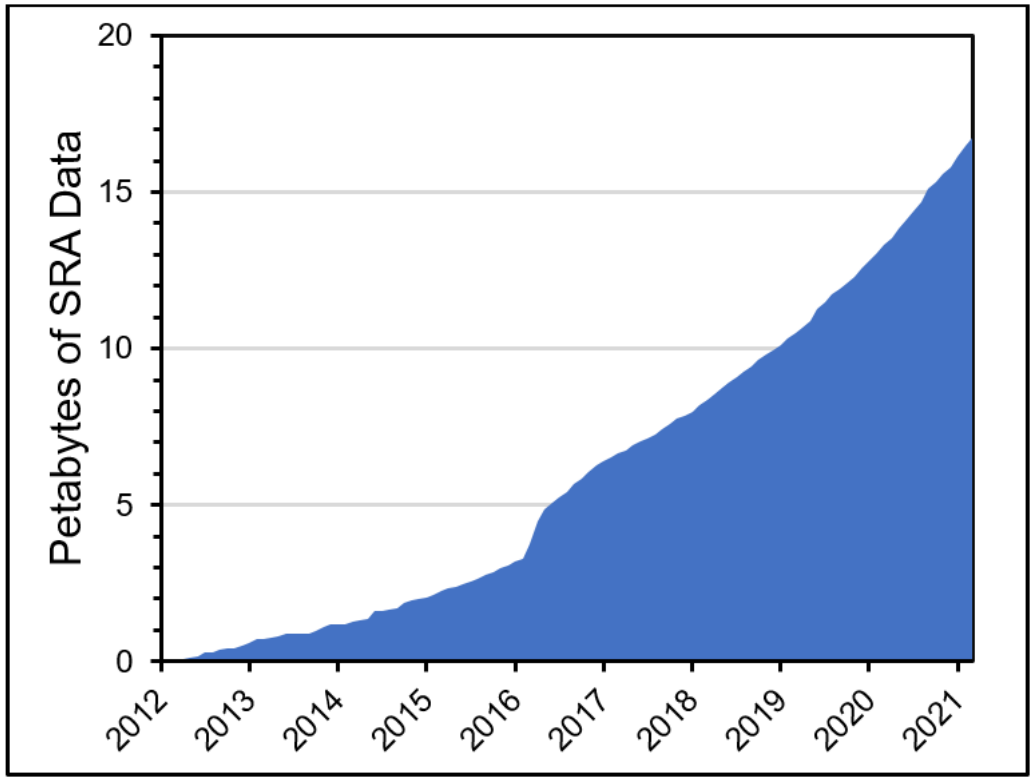
figshare

FAIRDOM HUB

GitHub

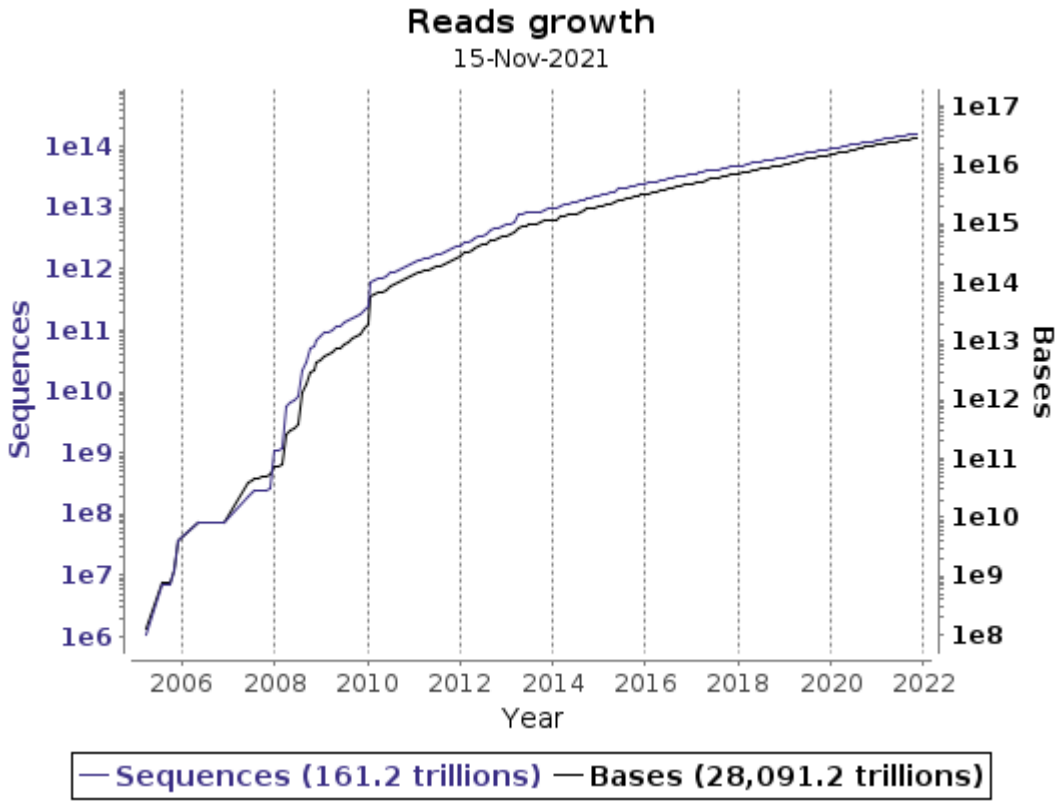
Podatkovni repozitoriji

Rast količine podatkov visokozmogljivostnega sekvenciranja:



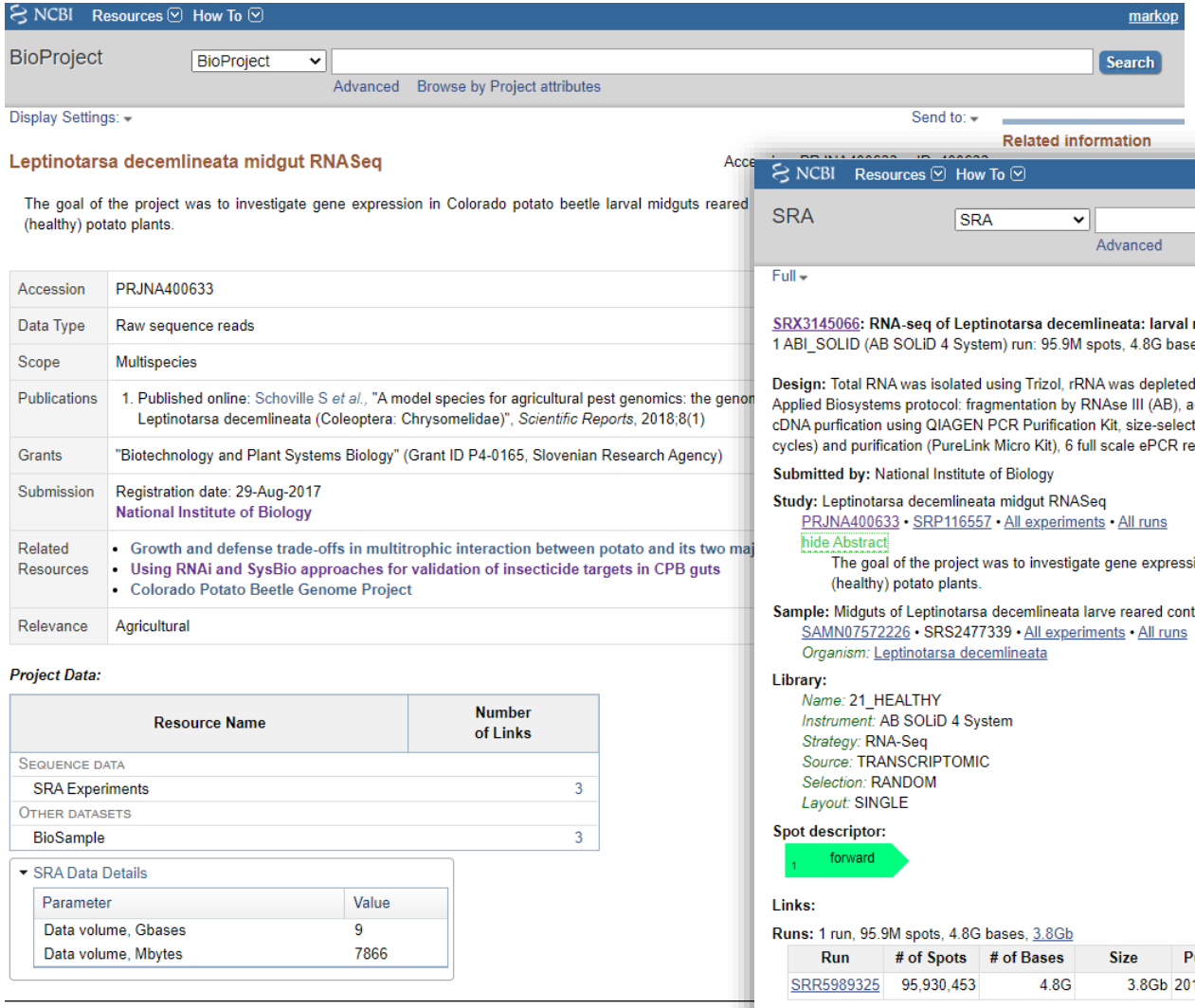
Vir: NCBI Insights
(https://ncbiinsights.ncbi.nlm.nih.gov/2021/08/09/espsss-workshop/#more-6180)

ENA



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

Deponiranje surovih podatkov v javne repozitorije



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

- 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…)

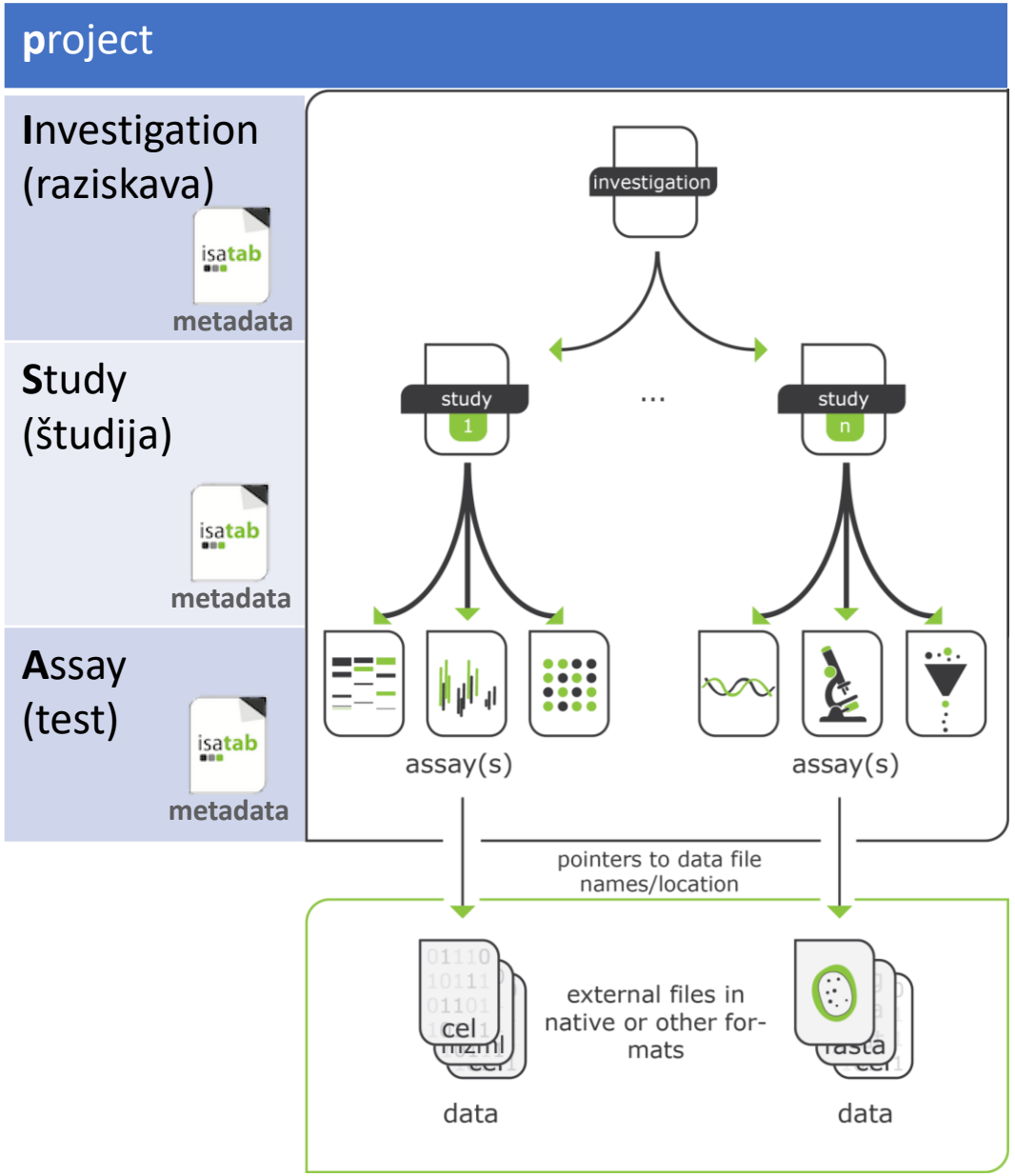
mrežni strežnik z varnostno kopijo

drugi mrežni strežnik z varnostno kopijo

pISA-tree

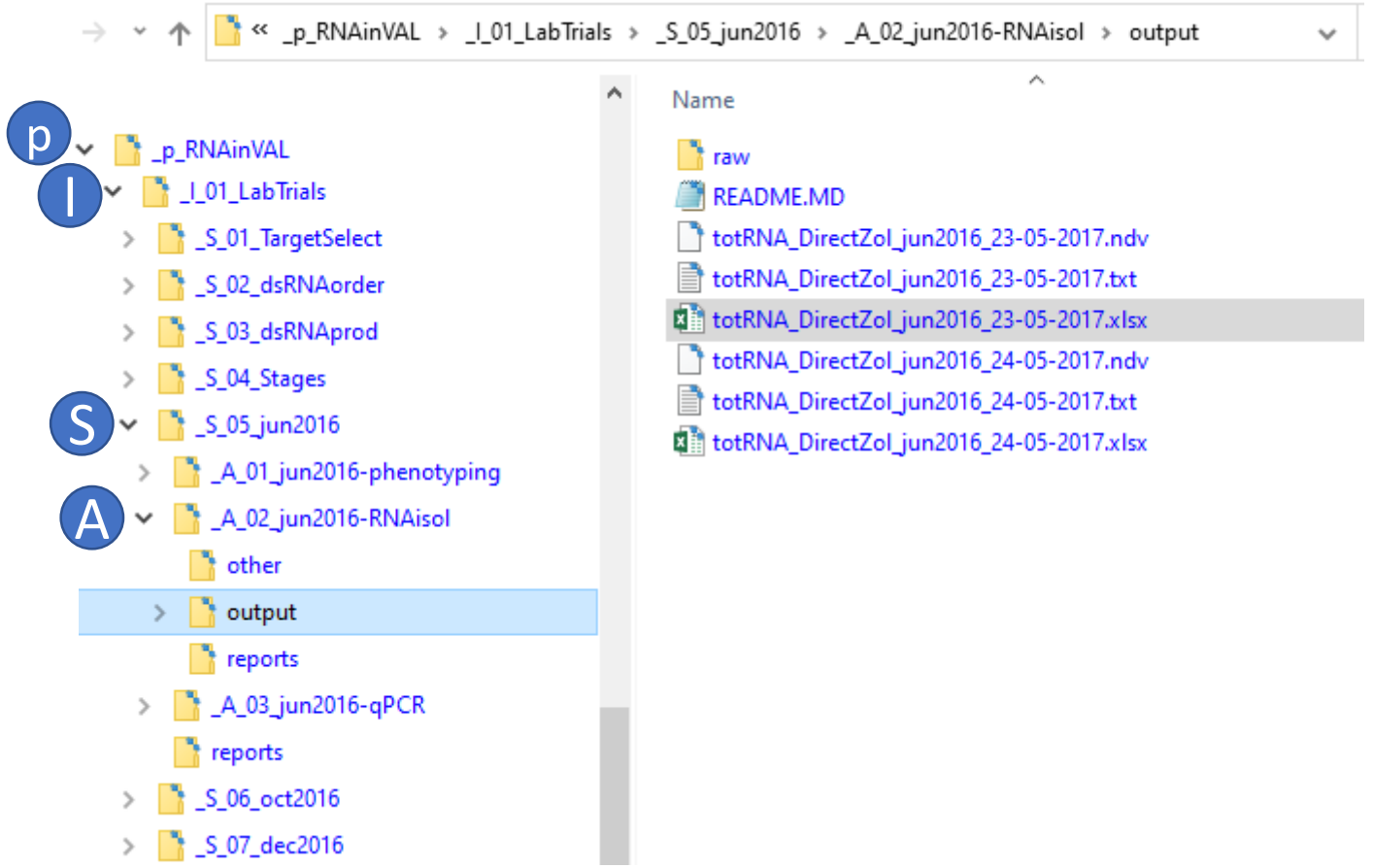
https://github.com/NIB-SI/pISA-tree

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

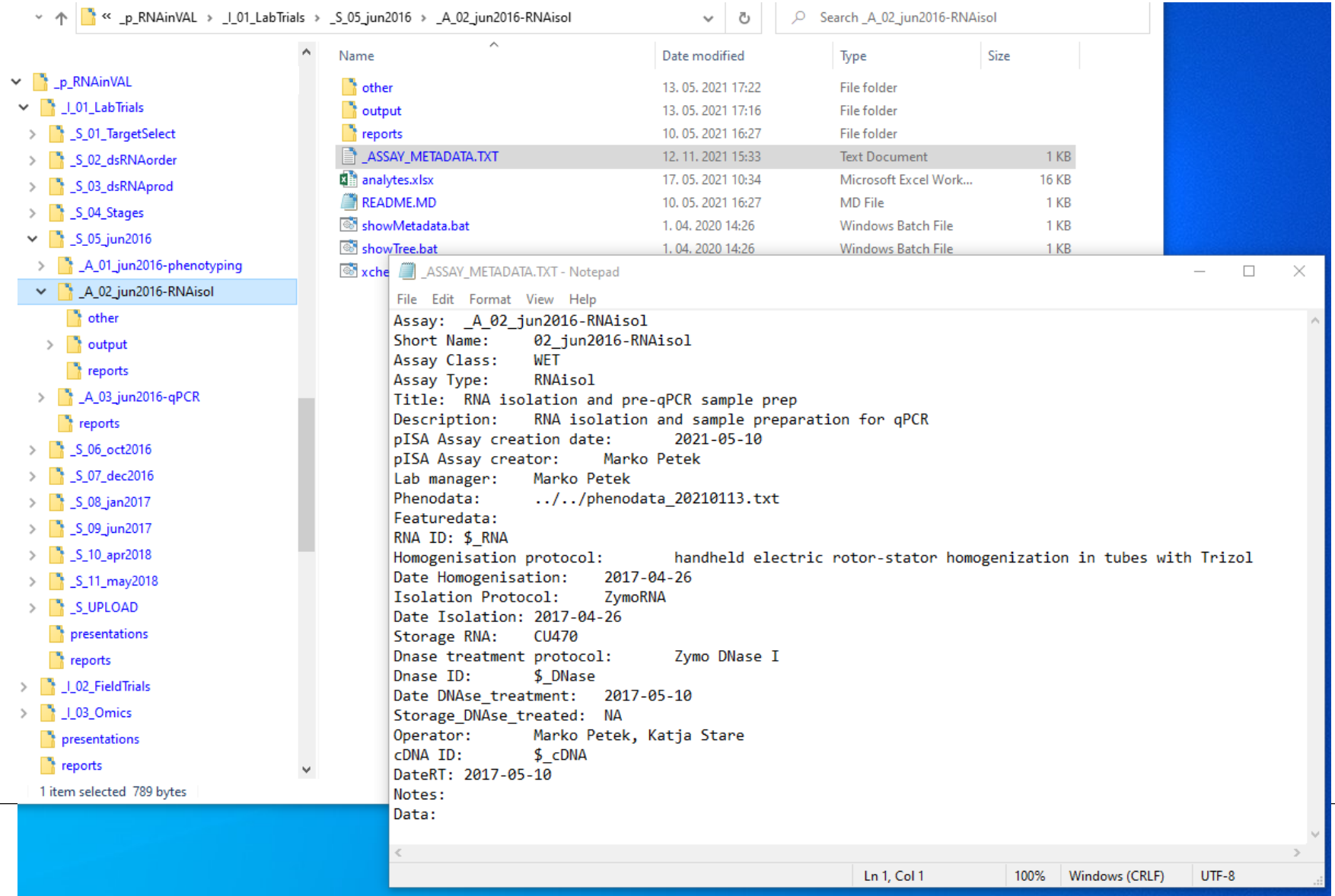


Vir: https://isa-tools.org/format/specification.html

pISA-tree: primer lokalne strukture



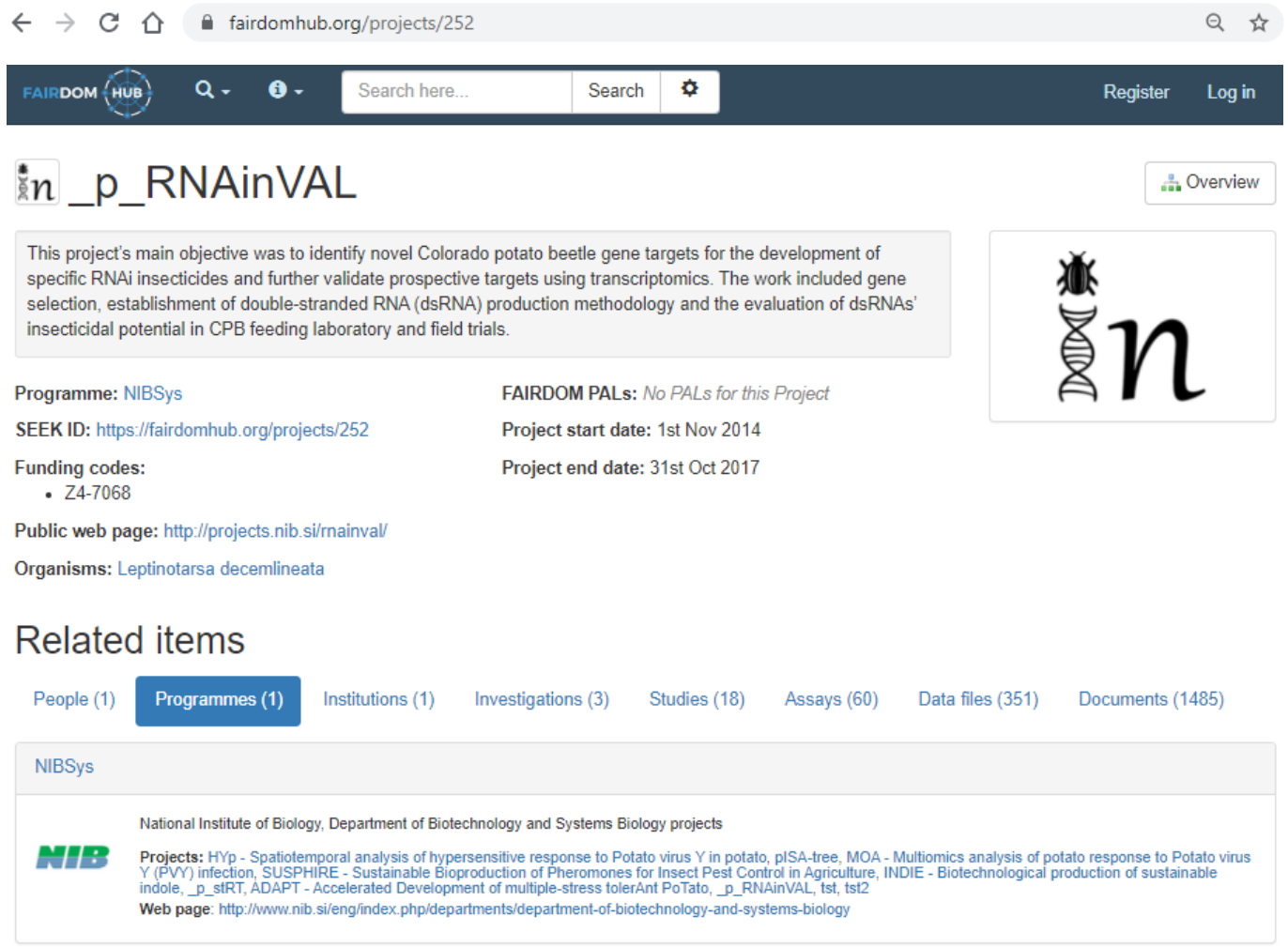
pISA-tree: primer lokalne strukture



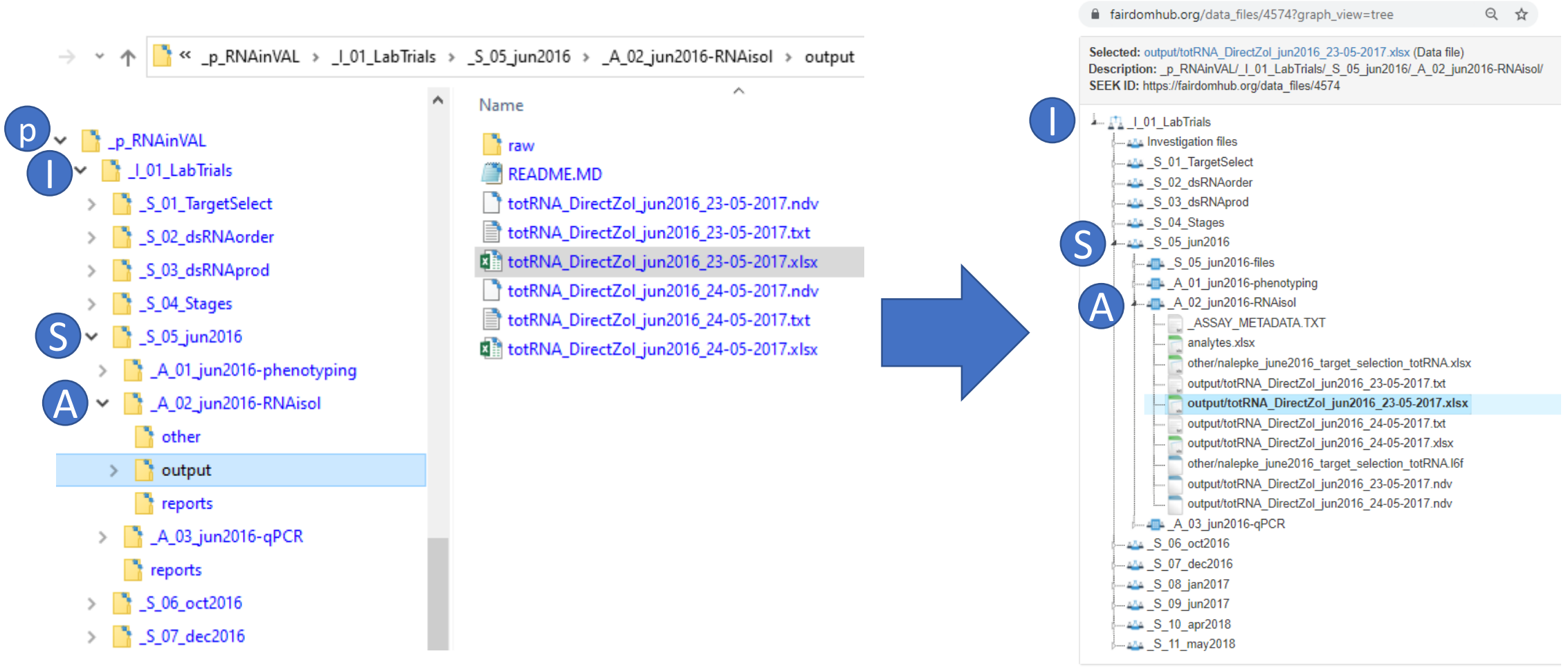
pISA-tree --> FAIRDOMHub.org

- 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)



pISA-tree ----------> FAIRDOMHub.org



Zahvala

prof. dr. Kristina Gruden

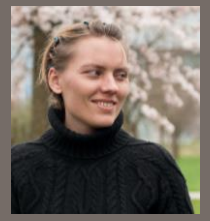
doc. dr. Špela Baebler

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