

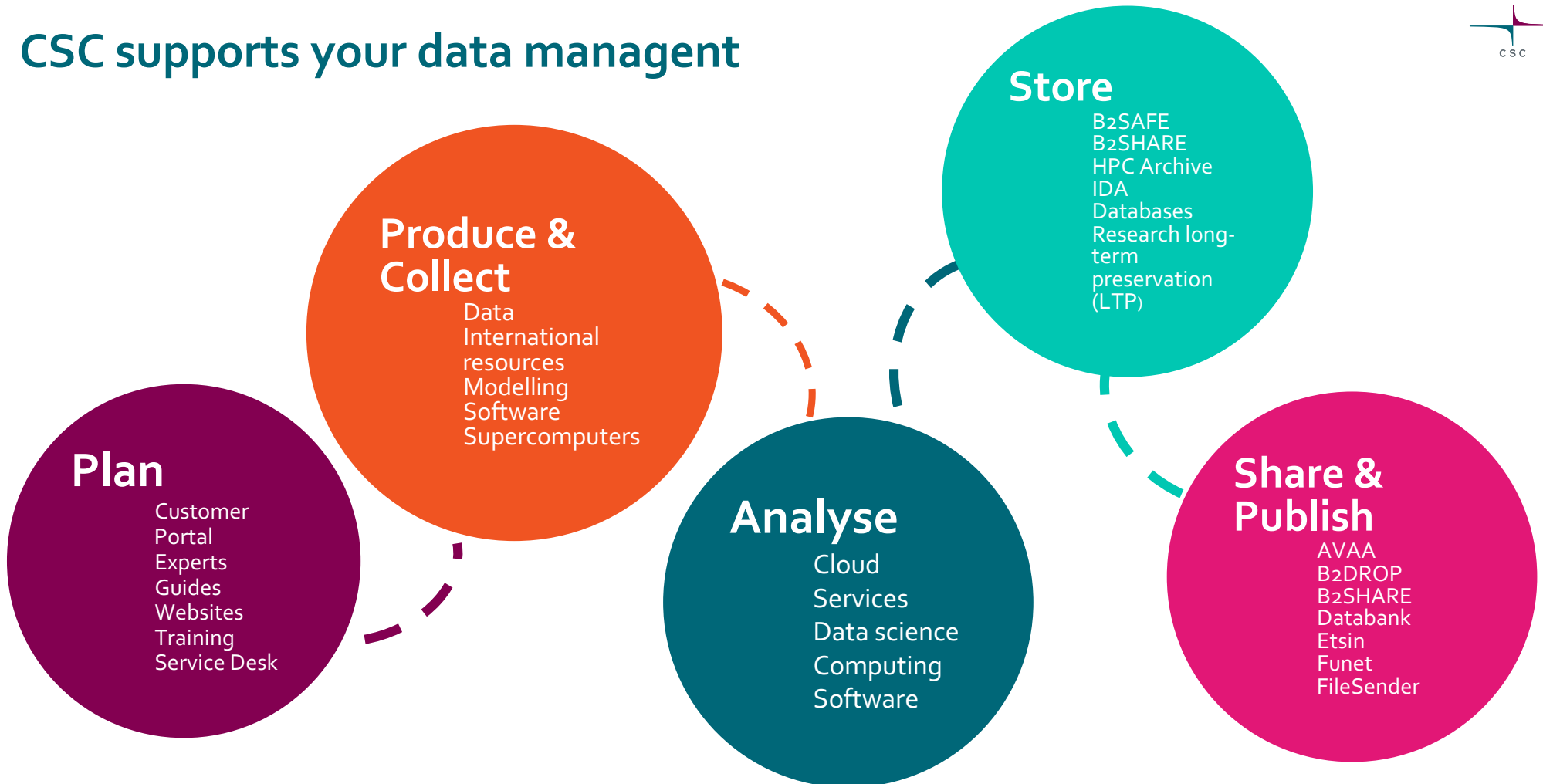
CSC services for digital humanities HELDIG Summit

Jessica PvE

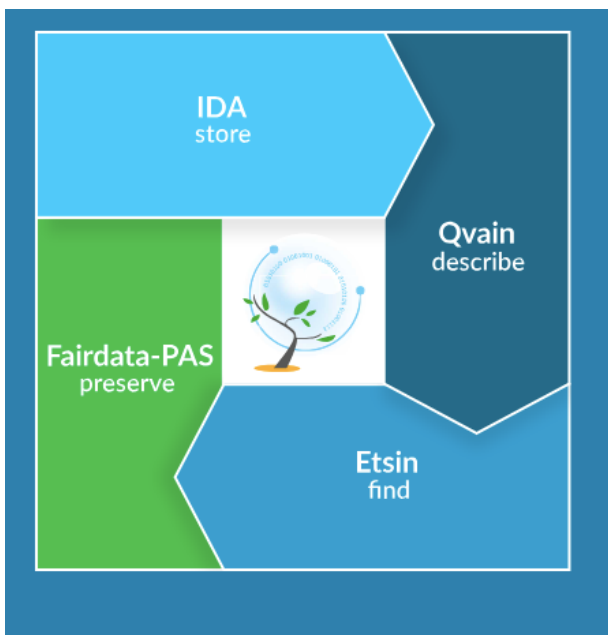






CSC – Suomalainen tutkimuksen, koulutuksen, kulttuurin ja julkishallinnon ICT-osaamiskeskus

CSC supports your data management



Findable **A**ccessible **I**nteroperable **R**eusable





FINDABLE
ACCESIBLE
INTEROPERABLE
REUSABLE



CITABLE

HOW TO GET THERE?

Roadmap  Stakeholders



Fairdata.fi



RESEARCH DATA

F

FINDABLE

- Described in relevant catalog with enough detail
- Landing page with globally unique identifier

A

ACCESSIBLE

- Can be retrieved over the internet
- Versioning and lifecycle documented
- Tombstone page if data is deleted

I

INTEROPERABLE

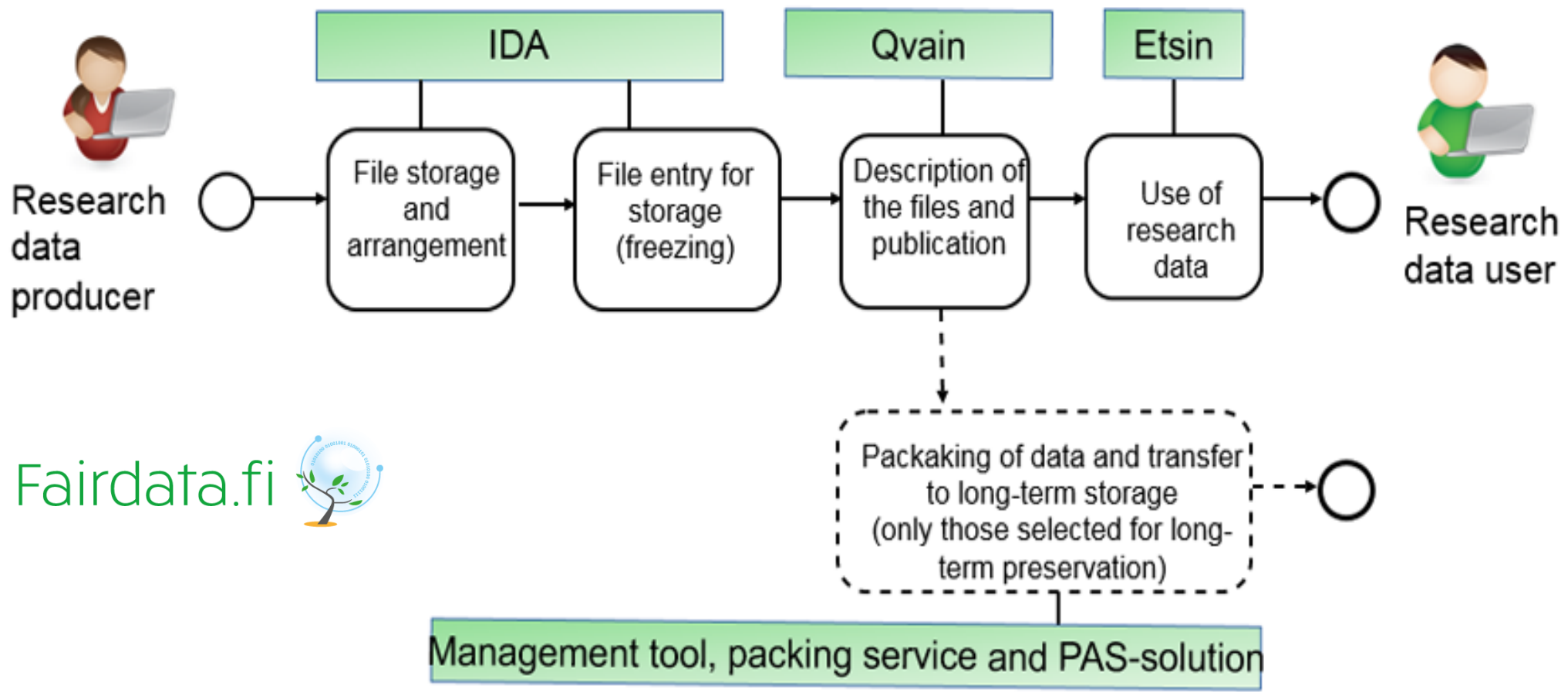
- Common, documented, and open formats

R

RE-USABLE

- Well documented and intelligible
- Rights clearly stated





Fairdata.fi



<https://hkm.finna.fi/Record/hkm.HKMS000005:km00000pnp>

Alanco Jan HKM, Valokuvaaja 02.08.1982

URN



B2DROP
Sync and Exchange Research Data

B2SHARE
Store and Share Research Data

B2SAFE
Replicate Research Data Safely

B2STAGE
Get Data to Computation

B2FIND
Find Research Data

B2HANDLE
Register your Research Data

B2ACCESS
Identity & Authorisation

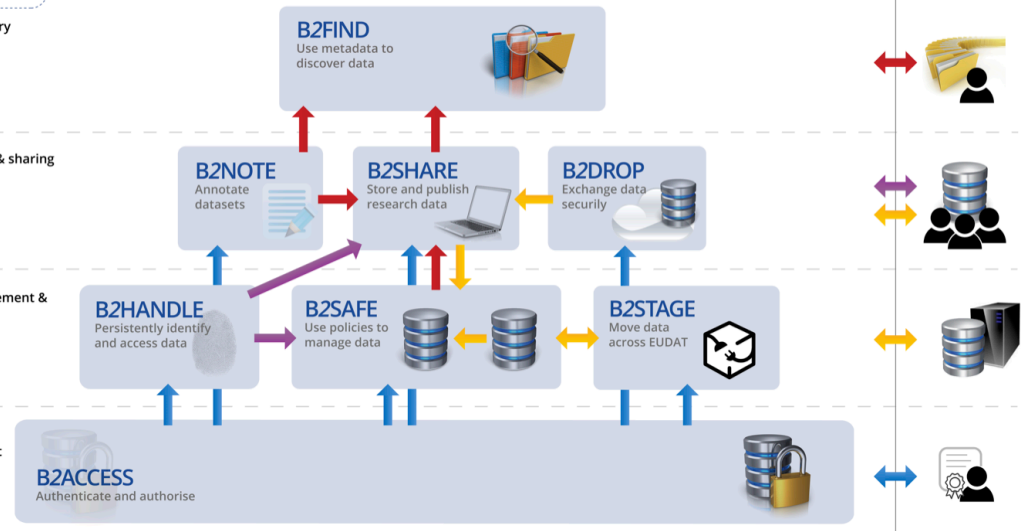


Data discovery

Data access & sharing

Data management & preservation

User management



Service Integration




 **B2SHARE**
Store and Share Research Data

 **B2DROP**
Sync and Exchange Research Data

Bidirectional
Integration


GO TO EUDAT WEBSITE

 B2SHARE

Select B2DROP files

File name	Size	Date
EUDAT2020		10/15/2015, 3:11:32 PM
EUDAT2020 KO 2015		4/2/2015, 4:01:08 PM
documents		9/5/2014, 10:39:14 PM
Conference Party		10/1/2014, 12:19:45 AM
EUDAT2 Graphics		2/22/2015, 11:38:27 PM
IRODS workshop		4/15/2015, 10:56:56 AM
<input type="checkbox"/> _DS_Store	6 KB	9/2/2015, 3:53:02 PM
<input type="checkbox"/> _DS_Store	4 KB	9/2/2015, 3:52:07 PM
<input type="checkbox"/> B2 Services integration.jpg	68 KB	9/2/2015, 3:52:52 PM
<input type="checkbox"/> ownCloudUserManual.pdf	2 MB	9/5/2014, 9:37:06 PM
<input type="checkbox"/> EUDAT-B2DROP.pdf	61 MB	9/17/2014, 1:08:21 PM

Cancel Select

Powered by 



EUDAT receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654065. Legal Notice.

GO TO EUDAT WEBSITE

Step 01









Drag and drop files here

Select files
Start upload

Filename	Size	Status	Remove
ELIXIR EBI Subscription Data Distribution use case v3.png	101 kb		

Step 02

Select a domain or project

- Linguistics 
- Herbarium 
- Biodiversity 
- Generic 
- Ontology 
- RDA 
- Hydro-Meteorology 
- Biomedical Research 

Deleted files

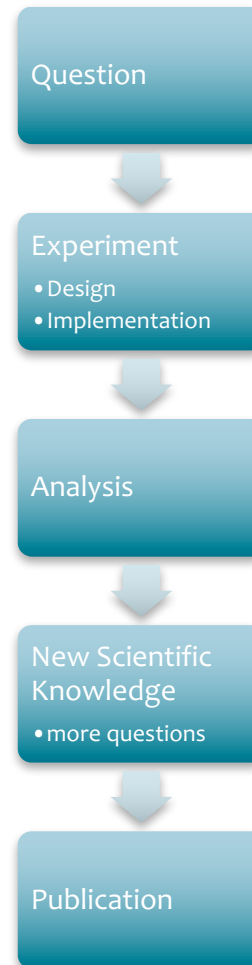
CSC's targets in data analytics

- Help customers reach higher levels of analytics maturity with **expert consultation**
- Provide open and tailored **cloud services** for data-driven solutions
- Support the Finnish society in digitalisation and artificial intelligence through **networking and co-creation**



CSC Expert support

Statistics
Data Science
Machine Learning
Neural Networks
Data Visualization



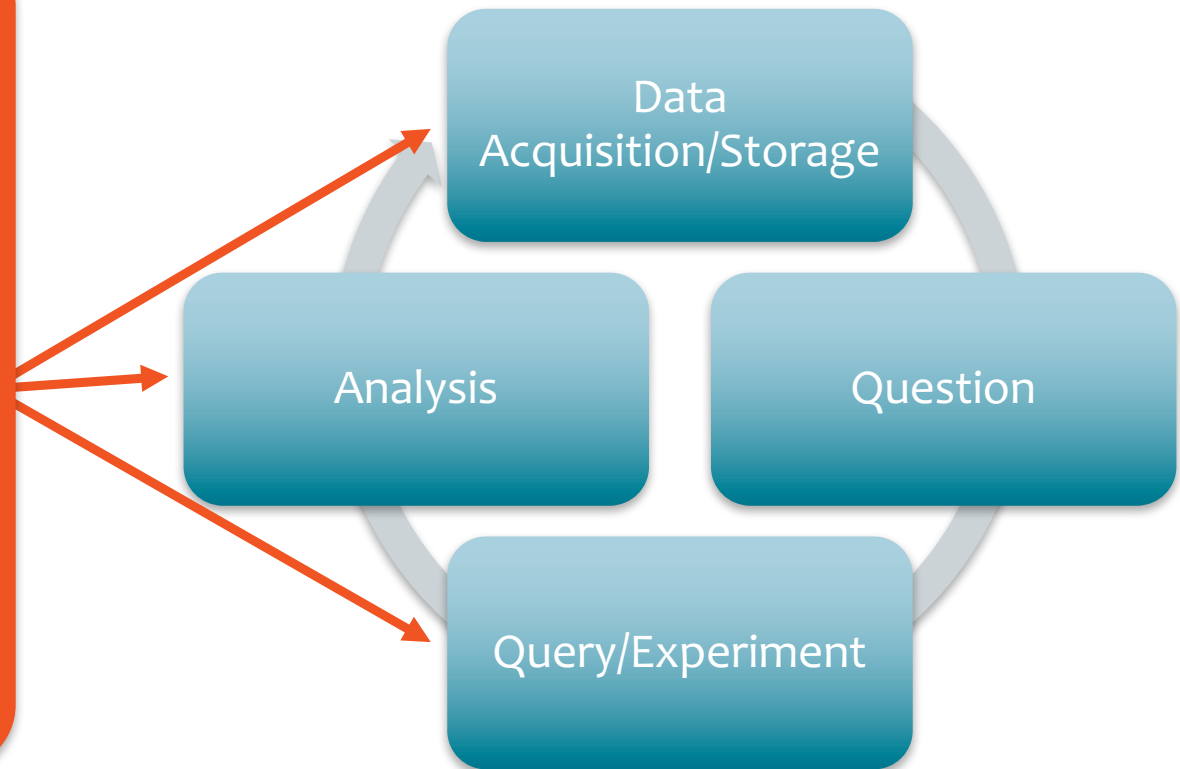
CSC Platform for big data workflows

Cloud platforms

- cPouta IaaS cloud
- Pouta Object Store
- ePouta IaaS cloud for **sensitive data**
- Rahti container cloud
- Coming: Rahti for **sensitive data**
- Elastic Spark for **big data**

HPC platforms

- Taito **GPU cluster**
- Coming: Data lake



High-Performance Digitisation project

- Develop intelligent machine learning based annotation pipelines for archive material (scanned newspapers, books and official documents)
 - Implement image recognition, image labeling, text labeling and feature detection
 - Efficient implementation with HPC and GPU's
- Funded by INEA/EU through CEF instrument (HPC & Open Data call)
- Collaboration with The National Library and The National Archives

CSC is your eInfrastructure partner

- CSC is part of many different communities that provide scientific eInfrastructure in language research, bioinformatics, genomics, environmental sciences, atmospheric sciences...
- Our computing environment is designed for tailoring and domain specific adaptations
- Happy to join and support communities that create novel scientific eInfrastructures together

Chipster

<http://chipster.csc.fi/>

→ Mylly

tailored for the Language Bank of Finland

<https://www.kielipankki.fi/support/mylly>



Some use cases for Mylly

1. Automatic speech recognition
2. Tokenize text and count word form frequencies
3. Automatic dependency parsing of Finnish text
4. Search from an annotated corpus via Korp API and get the results in a table (spreadsheet)





<https://research.csc.fi/>



Data storage



**Data management
planning and policies**



Data analytics



**Data sharing and open
data**



Data discovery



**Best practices and
guidance**



Jessica Parland-von Essen

Parland@csc.fi
@jpve



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc---it-center-for-science



github.com/CSCfi