CSC experiences on sharing HEI data in Finland

Heini Manninen
Development Manager
Reporting, Analytics and Insight Services for Education

BI SIG Winter Seminar
3rd February 2020

CSC
ICT Solutions for Brilliant Minds

Agenda

- CSC in short

- National infrastructure of educational and research information in Finland

  Case **vipunen** EDUCATION STATISTICS FINLAND

- Reasoning and value of setting up the national infrastructure

Heini Manninen

MSc (Econ) 2004 | Helsinki School of Economics (today Aalto University, School of Business)

taloustutkimus

B2B and B2C Market Research
Customer Experience
Quolitative Research

2004-2010

KANTAR TNS

B2B Market Research
Customer Experience
Employee Experience

2010-2011

Customer data management & analytics
BI & Strategy
Consumer Insight & Engagement
Segmentations
Business Cases & Game development

VEIKKAUS

2011-2018

Consumer Insight & Engagement
KPI's
Application development
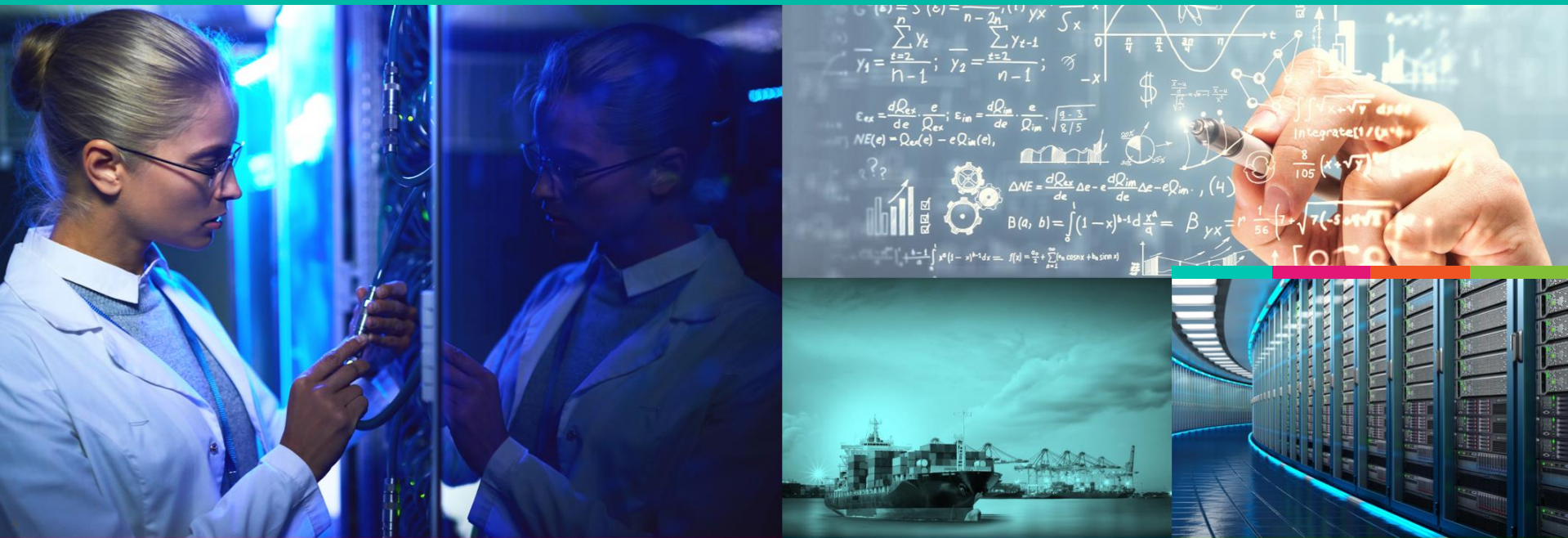Agile Digital Service Development

S-Pankki

2018-2019

CSC

Vipunen Project Manager 2019
Development Manager 2020 –

**Top team of 10
Developers, Application, Reporting,
and Analytics Specialists &
Coordinators**

Reporting, Analytics & Insight
Services for Education (RAISE)
BU: Services for Digitalization of Education

CSC in short

Non-profit state organization with special tasks

Turnover in 2018

44,9 M€

CSC - IT Center for Science Ltd.

Headquarters in Espoo
Data center in Kajaani

Owned by state 70% and all Finnish higher education institutions **30 %**

Circa

+400 employees in 2019

We provide high quality ICT expert services for

- **Research**
- **Education**
- **Culture**
- **Public administration**
- **Enterprises**

Our targets

- **Enable world-class data management and computing**
- **Maximize value of data**
- **Leverage artificial intelligence**
- **Pave way for lifelong learning**

Business Units

Services for Computational Research

Services for Digitalization of Education

Services for Data Management

+ Platform Services

National infrastructure of educational and research information in Finland
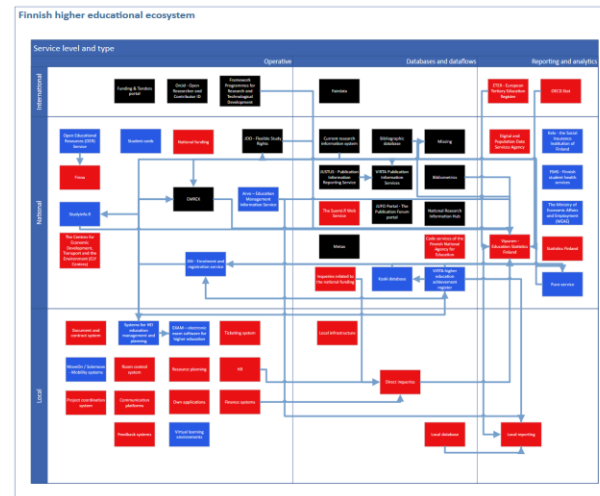
Management of national education information infrastructure

- The Ministry of Education and Culture (MoEC) and the Finnish National Agency for Education (EDUFI)
  - o Responsible for managing the national data warehouse and national data content

- HEIs
  - o Own internal data warehouse systems and infrastructures to support their performance
  - o Platforms, systems, practices and capabilities vary

- CSC
  - o Operates as a national service provider
  - o Is responsible for the technical realization of some national education data collection
  - o Offers information and data services and support for both of its owners according to their decision and priorities

Ecosystem of educational data in Finland

- To get the *big picture* on information and data flows between all operators can be tricky

- Working on a draft of ecosystem of educational data
  - Based on workshop with HEIs on September 2019

- Goals
  - To visualize the current ecosystem of HEI data sources and systems
  - To point out the relations between different systems and needs for development
 - To better support HEIs, MoEC and EDUFI in data and knowledge management

https://app.powerbi.com/view?r=eyJrIjoiZTk2NzVmNGItN2Q
zZS00YTc2LWJkMTctNjYxNTI0YjJjYmUxIiwidCI6IjkxMDczOD
lkLTQ0YjgtNDcxNio5ZGEyLWM0ZTNhY2YwMzBkYiIsImMiOj
h9

Vipunen background
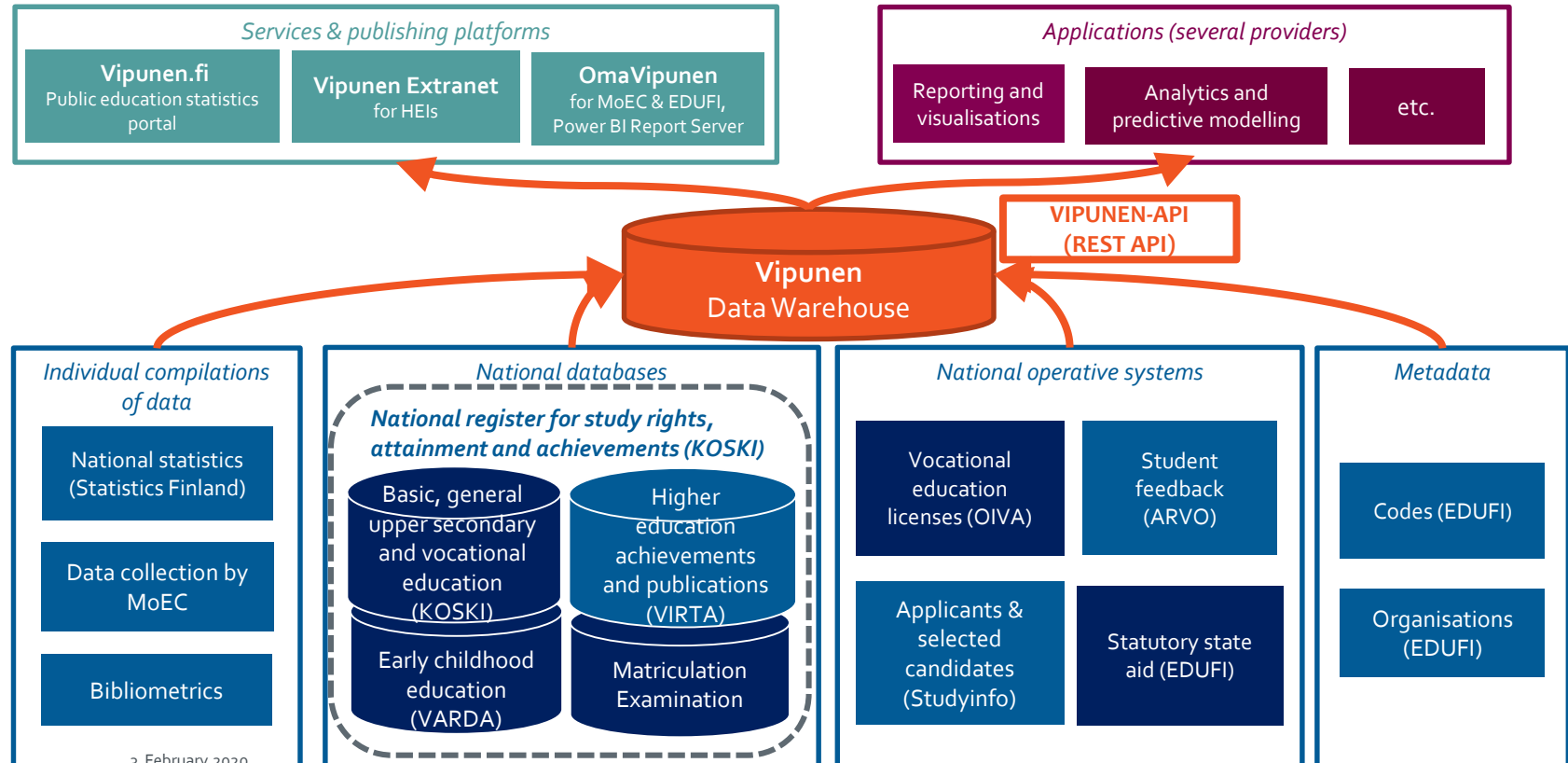## Information portal and data warehouse for education administration

- Early 2010's MoEC and EDUFI decided to build a reporting portal covering information on all levels of education

- Vipunen portal was released in 2013
  - Name *Vipunen* origins from the Finnish national epic the Kalevala
  - In Kalevala *Vipunen* is an old mythical giant and a wise man who has in possession valuable ancient knowledge

- At first Vipunen's information was based on Statistics Finland's data

- By 2016 Vipunen made available educational information from 6 different sources

- In 2018 Vipunen portal reached
  - over 120 000 user sessions
  - over 65 000 users
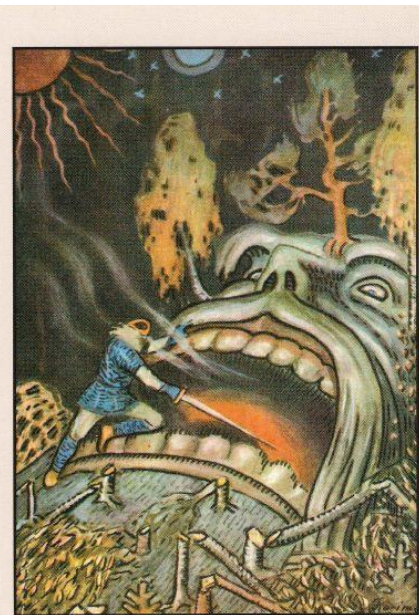
Virkkoi viisas Väinämöinen,
sekä virkkoi, että vaati:
"Ei sanat salahan joua,
eikä luottehet lovehen!"
Silloin virsikäs Vipunen,
Joll' oli suussa suuri tieto,
aukaisi sanaisen arkun,
virsilippahan levitti:
lauloi synnyt syltä myöten,
luottehet lomia myöten.

Vipunen Ecosystem today
Information portal and data warehouse for education administration

CSC

Services & publishing platforms

| Vipunen.fi Public education statistics portal | Vipunen Extranet for HEIs | OmaVipunen for MoEC & EDUFI, Power BI Report Server |

Applications (several providers)

| Reporting and visualisations | Analytics and predictive modelling | etc. |

Vipunen Data Warehouse

VIPUNEN-API (REST API)

Individual compilations of data

- National statistics (Statistics Finland)
- Data collection by MoEC
- Bibliometrics

National databases

National register for study rights, attainment and achievements (KOSKI)

- Basic, general upper secondary and vocational education (KOSKI)
- Higher education achievements and publications (VIRTA)
- Early childhood education (VARDA)
- Matriculation Examination

National operative systems

- Vocational education licenses (OIVA)
- Student feedback (ARVO)
- Applicants & selected candidates (Studyinfo)
- Statutory state aid (EDUFI)

Metadata

- Codes (EDUFI)
- Organisations (EDUFI)

Vipunen – Education Statistics Finland

- Public online portal: https://vipunen.fi/en-gb/

- All education levels represented

- Content is arranged according to the sectors of education
  - All levels and sectors
  - Pre-primary and basic education
  - General upper secondary education
  - Vocational education and training
  - University of applied sciences education
  - University education
  - Higher education and R&D activity
  - Population, educational and vocational structures

- Approx. 2 000 Excel reports and PBI data visulisations (Finnish, Swedish, English)

- Provides tools to explore data through various variables

3 February 2020

Vipunen Extranet

- Restricted extranet portal for higher education institutions

- More dynamic data and targeted reports for the HEIs

- Data sources
  - Higher education achievements and publications (VIRTA)
  - Data collection by the Ministry of Education and Culture
  - Applicants & selected candidates (Studyinfo/ EDUFI)
  - Student feedback (Arvo surveys)
  - National statistics (Statistics Finland)

- Layout and content mainly similar to public portal

- Same analysis tools as within the public Vipunen portal

vipunen KORKEAKOULUJEN EXTRANET

Data and reports

- Vipunen provides mainly
  - Excel Pivot reports
  - Power BI visualizations

- Vipunen API provides machine readable data for some data sets
  - Challenges e.g.: Data anonymity and demand to be open for everyone (no user authentication)

- MoEC and EDUFI have live access to Vipunen DW data cubes through Excel and Power BI
  - Availability of wider range of data variables and measures
  - Availability of un-anonymised data
  - Possibility to produce ad hoc and customized reports

- Let's look at some examples
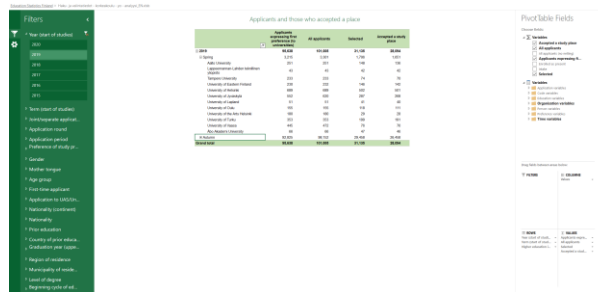
https://vipunen.fi/en-gb/_layouts/15/xlviewer.aspx?id=/en-gb/Reports/Haku-%20ja%20valintatiedot%20-%20korkeakoulu%20-%20yo%20-%20analyysi_EN.xlsb
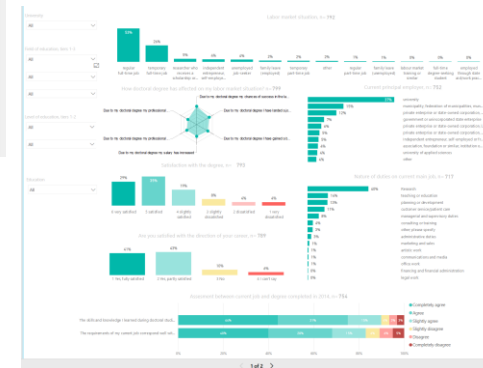
Applicants and those who accepted a place (universities)
Excel Pivot

https://vipunen.fi/en-gb/_layouts/15/xlviewer.aspx?id=/en-gb/Reports/Yliopistokoulutus%20-%20opiskelijapalaute%20-%20rahoitusmalli%20-%20vuosi_EN.xlsb

Student feedback for university at the graduation phase
Excel Pivot

Career Monitoring (doctoral degrees)
PBI Visualization

https://app.powerbi.com/view?r=eyJrIjoiNDVlZGM4N2EtNTg1Yy0oMDY4LThlMjAtYmEyZGVkNmY2MjY2IiwidCI6IjkxMDczODlkLTQoYjgtNDcxNio5ZGEyLWMoZTNhY2YwMzBkYiIsImMiOjh9

How HEIs use Vipunen and other information platforms
Source: Workshop with HEIs on Ecosystem of Educational Data, September 2019

CSC

How HEIs collect data?

- Excel files
- CSV files
- Application programming interfaces (APIs)
- Screen shots from browser
- Copying from browser

How HEIs present data forward?

- Power BI & Excel
- Data tables
- Graphics
- PowerPoint presentations
- Own reporting systems

To whom information is produced?

- Different levels of management (steering groups, faculties, departments, student affairs)
- Those responsible of operations development
- Communications and media relations
- Applicant marketing managers
- Lobbyists
- HR

## How information is used?

- Ancitipation of operations and funding
- Allocation of resources
- Applicant marketing development
- Career advice development
- Lobbying
- Recruiting

Most using mainly Excel or CSV

API used by only a few

Difficulties in combining public and organization's own data

Reasoning and value of setting up the national infrastructure
Challenges and development needs
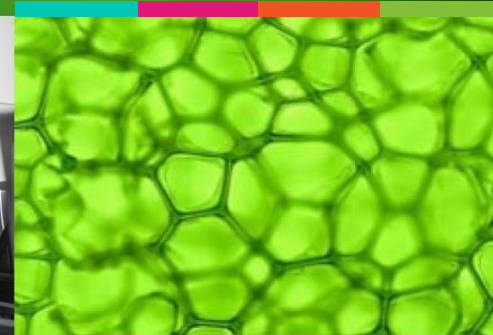
Reasoning and value of setting up national infrastructure

CSC

- Maximizing the value of data

- Making data available for various users

- Streamlining the use of data, e.g.:

- VIRTA-opintotietopalvelu (study information) provides higher education students data nationally to Finnish Student Health Services, Helsinki Region Transport and Social Insurance Institution of Finland (Kela)

- VIRTA-julkaisutietopalvelu (publications) is integrated to the Academy of Finland's funding system and lightens the administrative workload of researchers in reporting their publications

- Varda (early childhood education data system) provides up-to-date information on early childhood education to Kela, ministries, Finnish institute for health and welfare
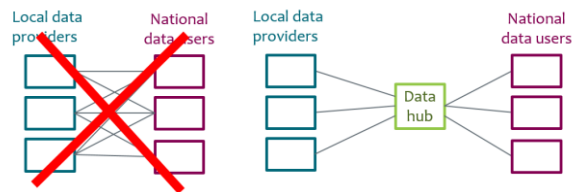
PROs for…

HEIs
✓ Possibility to check and evaluate data
✓ Benchmarks
✓ Lessen the need for local or own data systems

Ministry
✓ Structured data in common format
✓ Comparisons
✓ Possibility to check and evaluate data
✓ Leading with knowledge, eg. funding

Data hubs support the allocation of Ministry's national funding for education and research

Universities ~ 1,5 mrd €
- Degrees 30%
- Doctorals 8%
- Co-operation 1%
- Feedback 3%
- Career monitoring 2%
- Publications 14%
- Funding 12%
- Others 30%

Universities of Applied Sciences ~ 0,78 mrd €
- Degrees 56%
- Higher 6%
- Co-operation 1%
- Feedback 3%
- Career…
- Publications 2%
- Others 29%

Vocation schools ~ 1,7 mrd €
- Basic funding 50%
- Performance based funding 35%
- Impact 15%
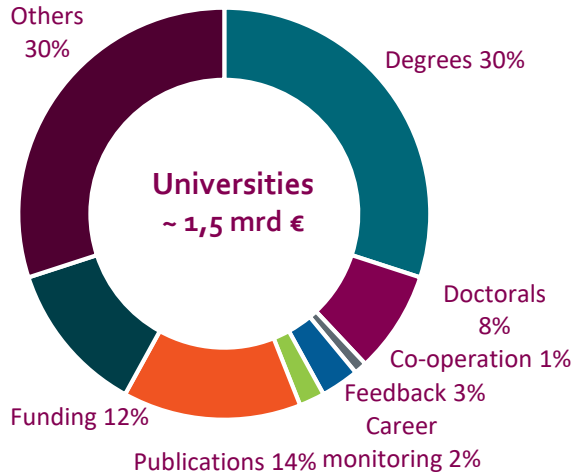
VIRTA-opintotietopalvelu, EMREX: 1 076 milj. €

VIRTA-publications, publication portal: 234 milj. €

KOTA, Ministry's own data surveys, KTLO-raportointi- ja analytiikka: 187 milj. €

ARVO – student feedback system: 125 milj. €

JUSTUS – publication records: 16 milj. €
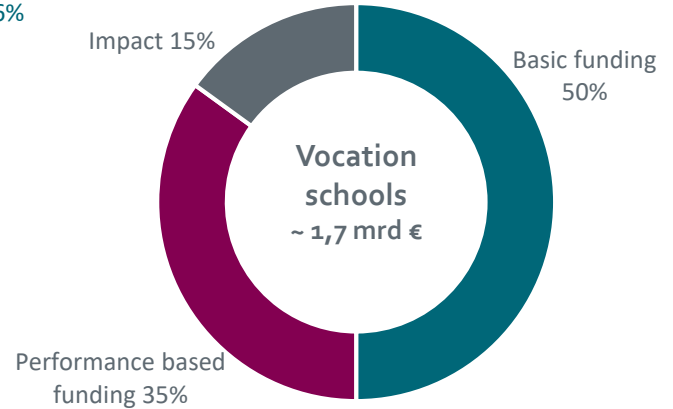
Vocational schools (OIVA, AMOS) steering system (funding)

Challenges & HEI's needs identified by CSC
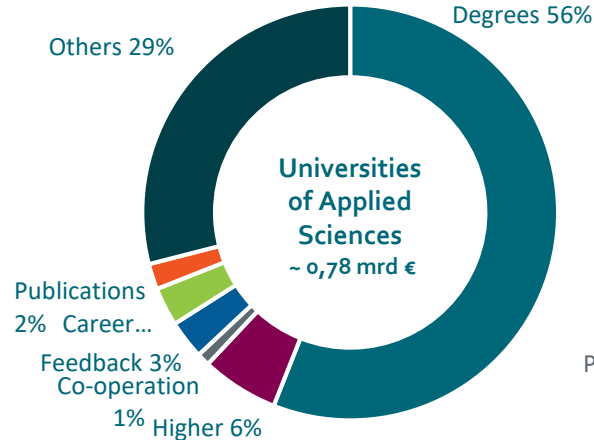## National infrastructure of educational and research information

- API demand is growing: HEIs are interested in row level "raw" data, self-service analytics and customized reporting

- Need for data that is structured by HEI's own organizational structure or other customized way

- Need to save customized Vipunen report templates for later use and updates

- Vipunen information portal's usability, management and maintenance
  - Vipunen portal and its data content and sources have grown vastly in the last years

- GDPR (General Data Protection Regulation)

- Some challenges in sharing data and abilities to decide on common practices and systems – many viewpoints and decision makers, sometimes even conflicting needs
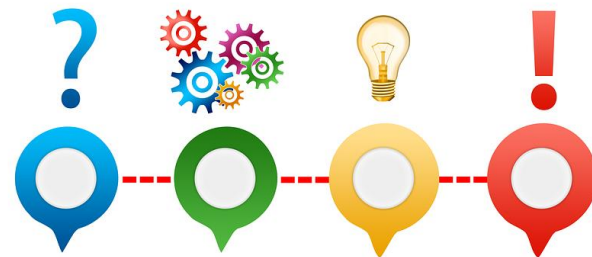
TOP 10 development needs by HEIs
Source: Workshop with HEIs on ecosystem of educational data, September 2019

Select 1-3 most important issues to develop

Rank	Issue
1st	Data that is structured by HEI's own organizational structure to Vipunen
2nd	Individual data instead of aggregations
3rd	VIRTA student information service user interface should be easier to use (not only machine readable data)
4th	Application objects and degrees should be linked
5th	Information and data related to "Digital Education Vision 2030" should be made available
6th	Automatic data from The Digital and Population Data Services Agency
7th	Time stamp to Vipunen data sets
8th	Possibility to follow research funding
9th	Statistic Finland's rankings
10th	ARVO feed back system's API development
11th	Statistic Finland's reports in RTF

23

CSC

Questions & discussion

Heini Manninen
Development Manager

heini.manninen@csc.fi
linkedin.com/in/heini-manninen-9157553
@HeiniManninen

Looking forward to connect

facebook.com/CSCfi

twitter.com/CSCfi

youtube.com/CSCfi

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

github.com/CSCfi

Vipunen System Architecture

Vipunen
Järjestelmä- ja teknologia-arkkitehtuurit

Architecture

Platform

Virtuaalikoneet - Windows + CentOS

Data base

Microsoft SQL Server, SSIS, T-SQL, SSAS, SSIS, Sharepoint

Version control

GitHub, Sharepoint

Technologies

DAX, Jenkins, SQL, C#, Python 2 +3, Java, java-tomcat (API), R, PDI, PowerShell, PowerBI

Maintenance

CSC

Data contents

- Arvo
- ETER
- OIVA
- ODW
- OPH-Koodisto
- OPH-Organisaatio
- VALOS
- VIRTA-JTP
- VIRTA-OTP
- YTL
- Suorat
- TK-H9098
- KOSKI
- Varda

Rest

DB

Data sources

ULKOISET KV-TIETOLÄHTEET

ETER (European Tertiary Education Register)
OECD
Bibliometriikka (Web of Science, Scopus)

OKM (MoEC)

OIVA (Opetushallinnon ohjaus- ja säätelypalvelu)
KOTA-tiedonkeruut
Arvo- opetushallinnon vaikuttavuustietopalvelu
VIRTA-julkaisutietopalvelu
VIRTA-opintotietopalvelu

OPH (EDUFI)

Organisaatiopalvelu
Koodistopalvelu
Koulutustarjonta
ePerusteet
Opiskelijavalintarekisteri
KOSKI-tietovaranto
VALOS (opetus- ja kulttuuritoimen rahoitusjärjestelmän rahoitus)
Varda varhaiskasvatuksen tietovaranto

Ylioppilastutkintolautakunta

Ylioppilastutkintorekisteri

Tilastokeskus (Stat. Fin)

TK-H9098 aineistot