



FRiDa

The Fair Research
Information Dashboard

A solution for the post CRIS era?



Julie Kirwan (analytics engineer)

Nick Veenstra (data engineer, research
information specialist)

Disclaimer / intro

- › Worked with Elsevier's Pure for 10 years at Eindhoven University of Technology (TU/e) and University of Groningen (RUG)
- › Kickstarted my career as research information specialist (coming from library software engineering)
- › Many useful developments e.g. CRUD api
- › However Research information landscape is changing fast:
 - . Data autonomy, AI, Business Intelligence

Topics

- Current CRIS situation*
- What is FRIDa?
- What the data tells us
- AI impact
- Do we still need a swiss army knife CRIS system?



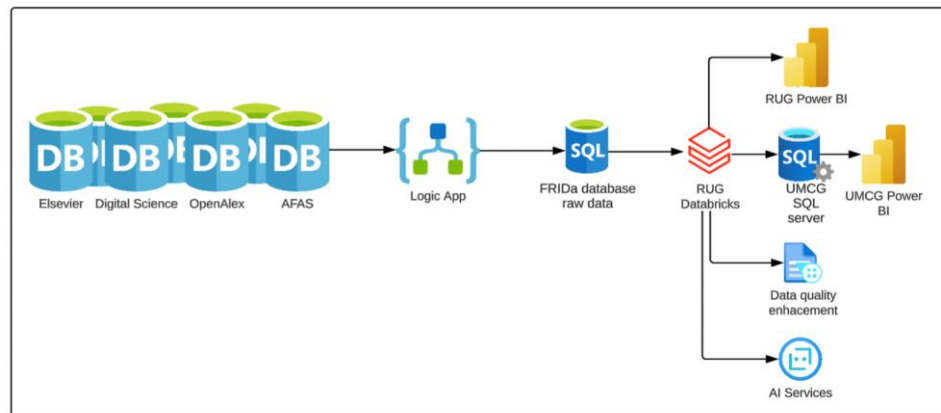
*: based on Pure experience

Current CRIS situation in NL

- Almost all universities use Pure
 - 10 year mark, tenders coming up
- 10 year review:
 - Apparently we do not need repositories
 - Elsevier as a software company
 - Software creep / roadmap vs dev capacity
 - Reporting module, portal
 - Funding opportunities
 - Regional priorities
- Focus on data autonomy (RUG)
- Advent of Research Information BI

What is FRIDa?

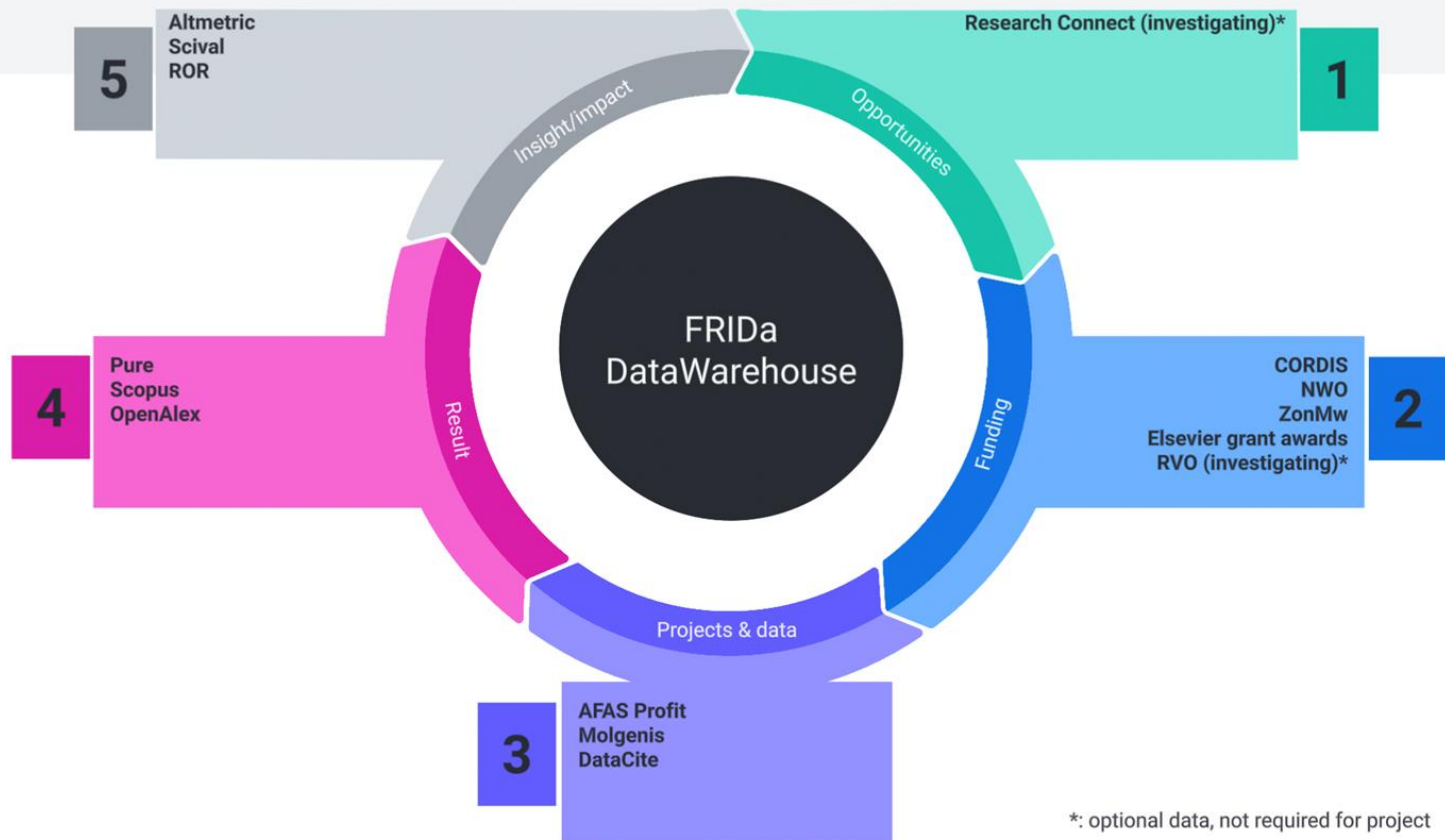
- › Cloud (Azure) based BI
- › Collects and tracks data from (currently) 13 sources
- › Creates a RUG “single source of truth” data warehouse
- › Initial users: management, policy advisers, funding officers and their support



FRIDa contents

What's in FRIDa?

March 2025



CRIS issues

→ “Current Research”

- ◆ Friction between CV and accounting use cases

→ Takes long to get reporting and portal right (AI will take over)

→ Dev capacity priority elsewhere?

→ Manual curation becoming impossible

- ◆ Major data quality issues
- ◆ Validation / support staff FTE shrinking

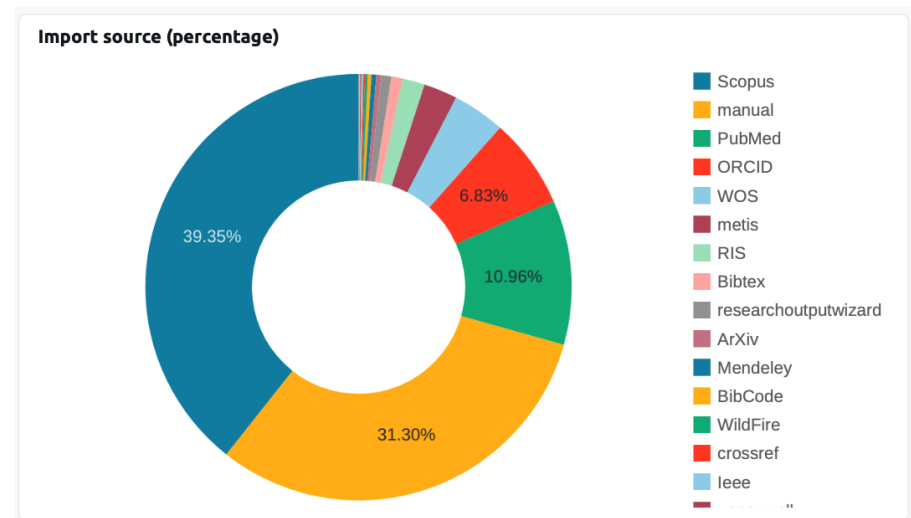
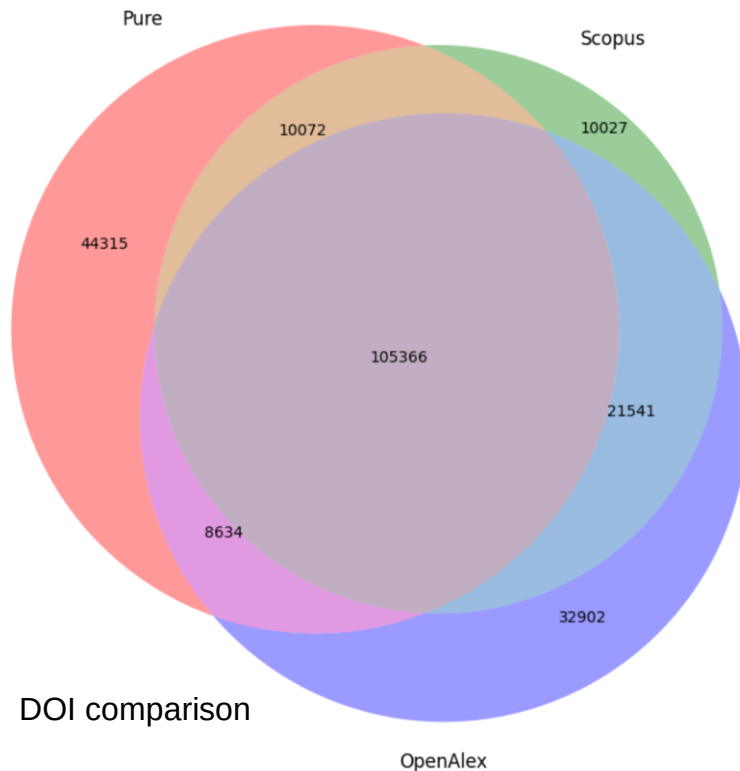
→ No middleware / architecture vision

- ◆ Multiple source systems for a datatype (increases workload)
- ◆ Organization affiliation matrix vs tree

→ CV not portable to new employer = less incentive to maintain

What FRIDa data already tells us

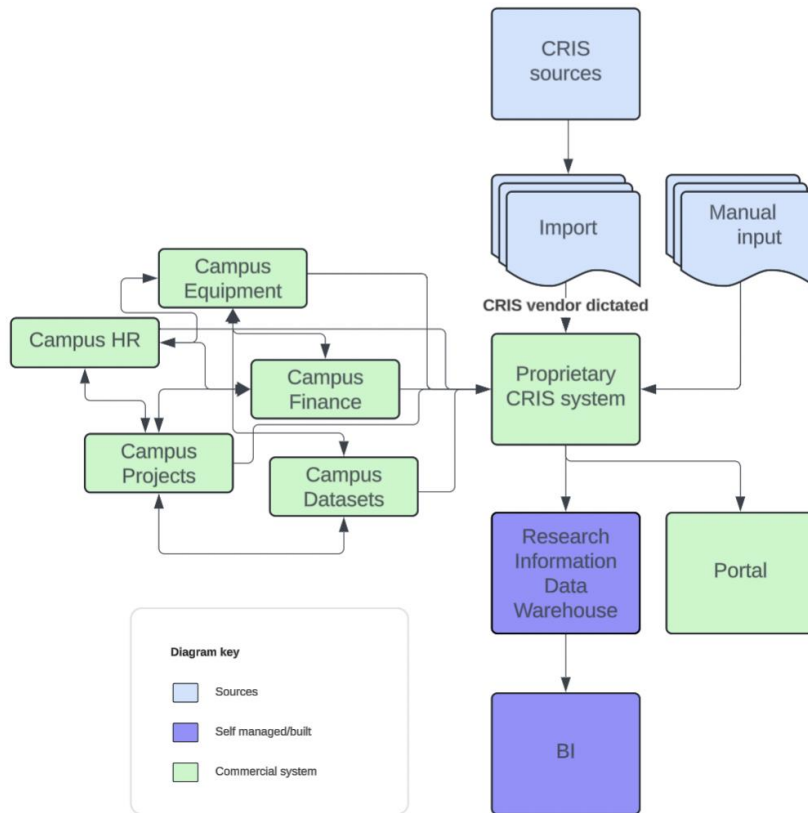
- › 70% of data is out there (can be imported)
- › Structuring to a base set is a challenge



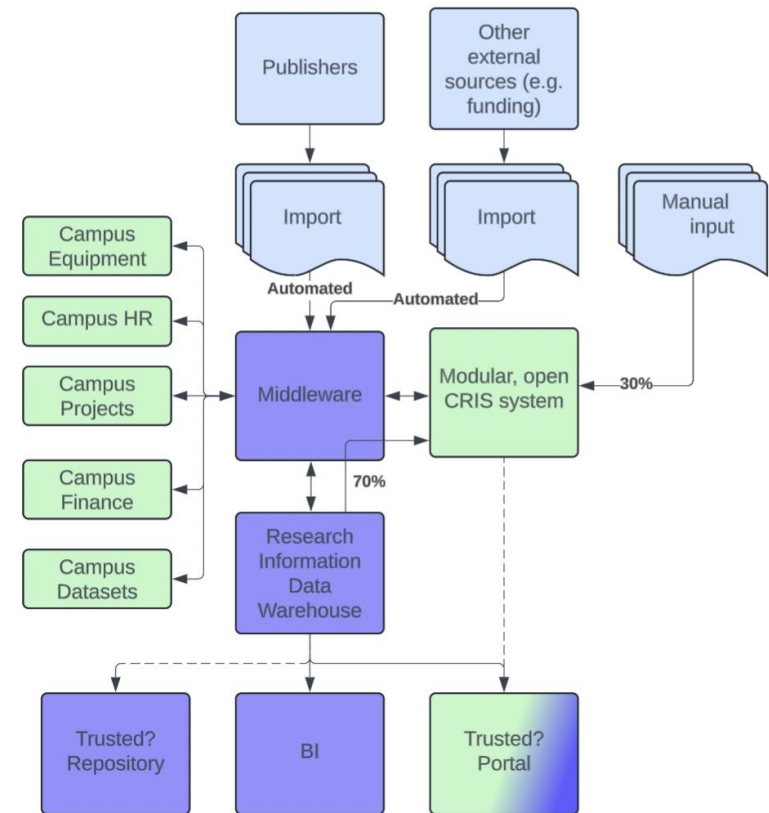
Pure publication source id

Data autonomy / architecture

Current



Desired



AI impact

- › *Chat support:*
 - › Immensely sped up FRIDa development
 - › Easier to BYO data warehouse
- › *AI agents:*
 - Load data from import sources
 - Assess import quality -> signal validation team/user
- › *AI function calling:*
 - Data quality tasks (classification, cleaning, deduplication)
- › *Retrieval Augmented Generation (RAG)*
 - › Pure data => CRIS Bot

So what do we need?

- CRIS => Research CV System: researcher first
- Primarily geared towards data entry / CV management of the “missing” part (our 30%)
- Loads from dwh (70%) to create complete user CV
- Optional reporting/accounting (not needed when doing BI, AI)
- Portal will become less relevant due to AI (also harvesting issues)
- Uses AI for workload leverage, showcasing
- Autonomy: updated, future proof flexible and open data model + API
- Open record/CV exchange format (Cerif?)

STAMFORD, Conn., February 19, 2024



Gartner Predicts Search Engine Volume Will Drop 25% by 2026, Due to AI Chatbots and Other Virtual Agents

Thanks for listening!
Questions?