


Researcher, LIS dept, Charles University, Prague CZ
CRIS analyst & maintainer, Czech Tech Univ, Prague CZ
CERIF and CRIS Architectures Task Group leader, euroCRIS
Jan Dvořák

ORCID IDS & ADMIN IDS

 <https://orcid.org/0000-0001-8985-152X>
vedik: 222221
Jan.Dvorak@ff.cuni.cz



The Outline of my Story

- When two PID systems have the same referents, we can study how well they agree and where they disagree

It's like the man with two watches!

- This situation is found in the national CRIS for Czechia
Two PID systems for researchers:
 - (1) ORCID iDs
 - (2) administrative IDs
- What we can see from the data so far
- How would YOU suggest the story should CONTINUE?

IS VaVal — the national CRIS for Czechia

CRIS = Current Research Information System, a.k.a. → [euroCRIS.org](https://eurocris.org)
RIMS = Research Information Management System

- Tracks public funding for research, technology development and innovation
- Institutions report their research outputs (incl. the links to the fundings)
- Basis for research evaluation exercises
- Data since 1994
- Currently at <https://www.rvvi.cz/>
Alas, no English version of the portal currently

The administrative IDs

- “The Native Number” (Czech: *rodné číslo*) – assigned:
 - *to citizens of the Czech Republic*
 - *to persons with residence permit in the country*
- 10-digit numbers (incl. one check digit)
- e.g. 700222/1234
- Encode the date of birth, the sex & the region of birth

The administrative IDs

+ They work – They must not be exposed (personal data protection)

They have traditionally been collected in the Czech national CRIS

Since 2014 they are transformed and exposed



„Researcher identification code“ (Czech: *vědíík*)

Pseudonymized numbers (randomly assigned)

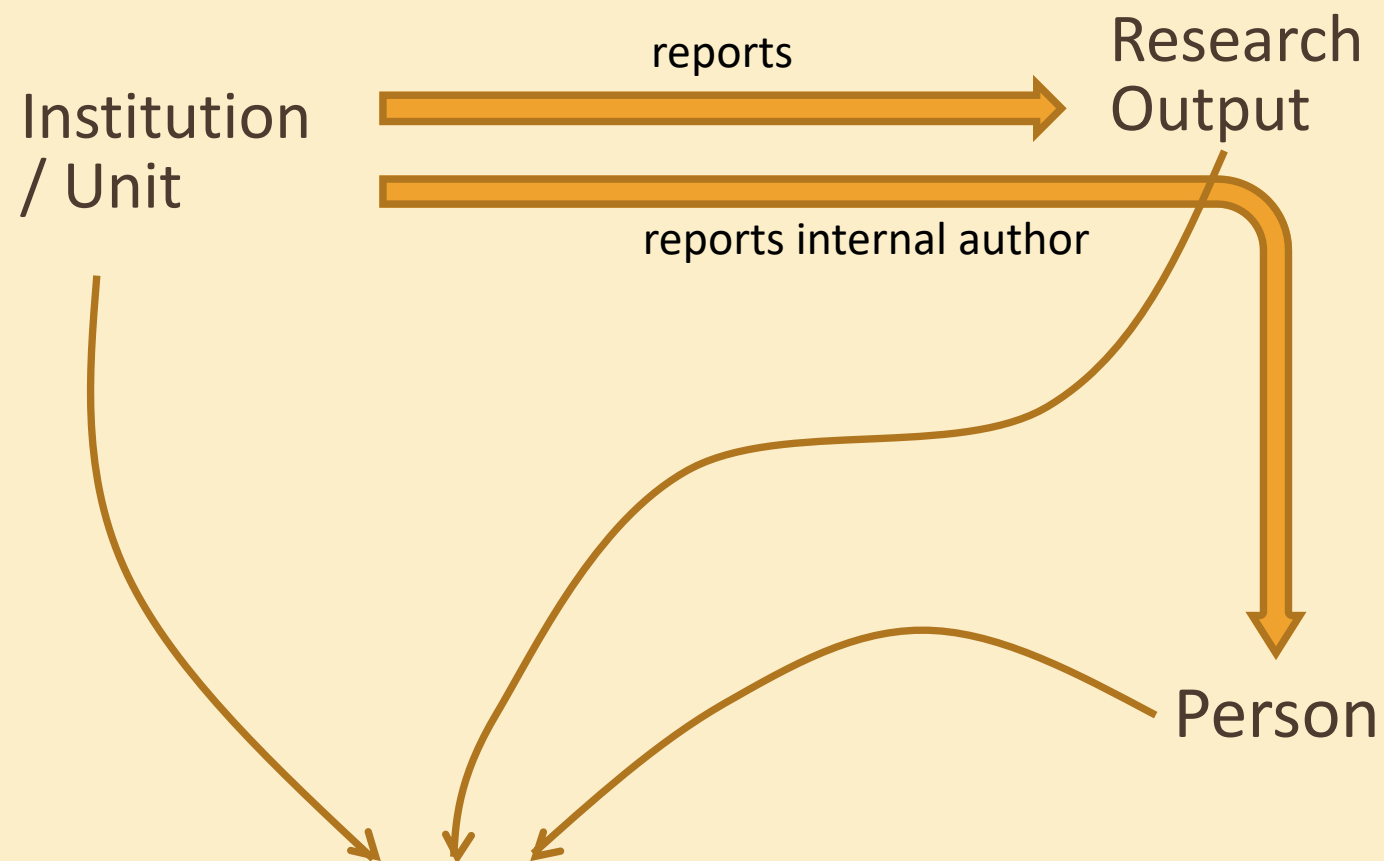
7-digits (incl. one check digit)

e.g. 2222221

ORCID iDs in IS VaVal

- Collected as optional since 2018
- => now 2 years of data

What is reported in the CRIS?



=> Individual affiliations of publication authors from the institution

An example: Article „*Search for charged Higgs bosons decaying into top and bottom quarks at root $s=13\text{TeV}$ with the ATLAS detector*“ (2018, J High Energy Physics, DOI [http://dx.doi.org/10.1007/JHEP11\(2018\)085](http://dx.doi.org/10.1007/JHEP11(2018)085)) as reported by the Czech Technical University in Prague / Faculty of Nuclear Sciences and Physical Engineering

Total authors → 2921

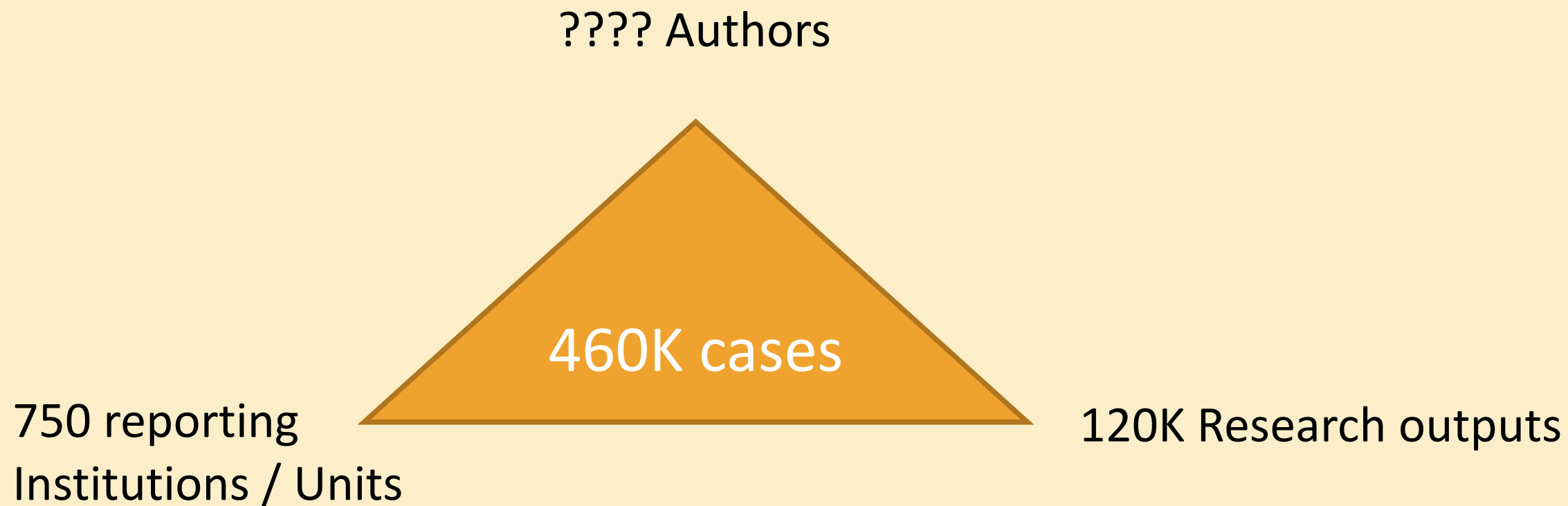
„Our“ authors → 9

8 our authors reported with the admin ID →
6 our authors reported with the ORCID iD →

6 authors from the same institution, different faculty, reported with the ORCID iD →

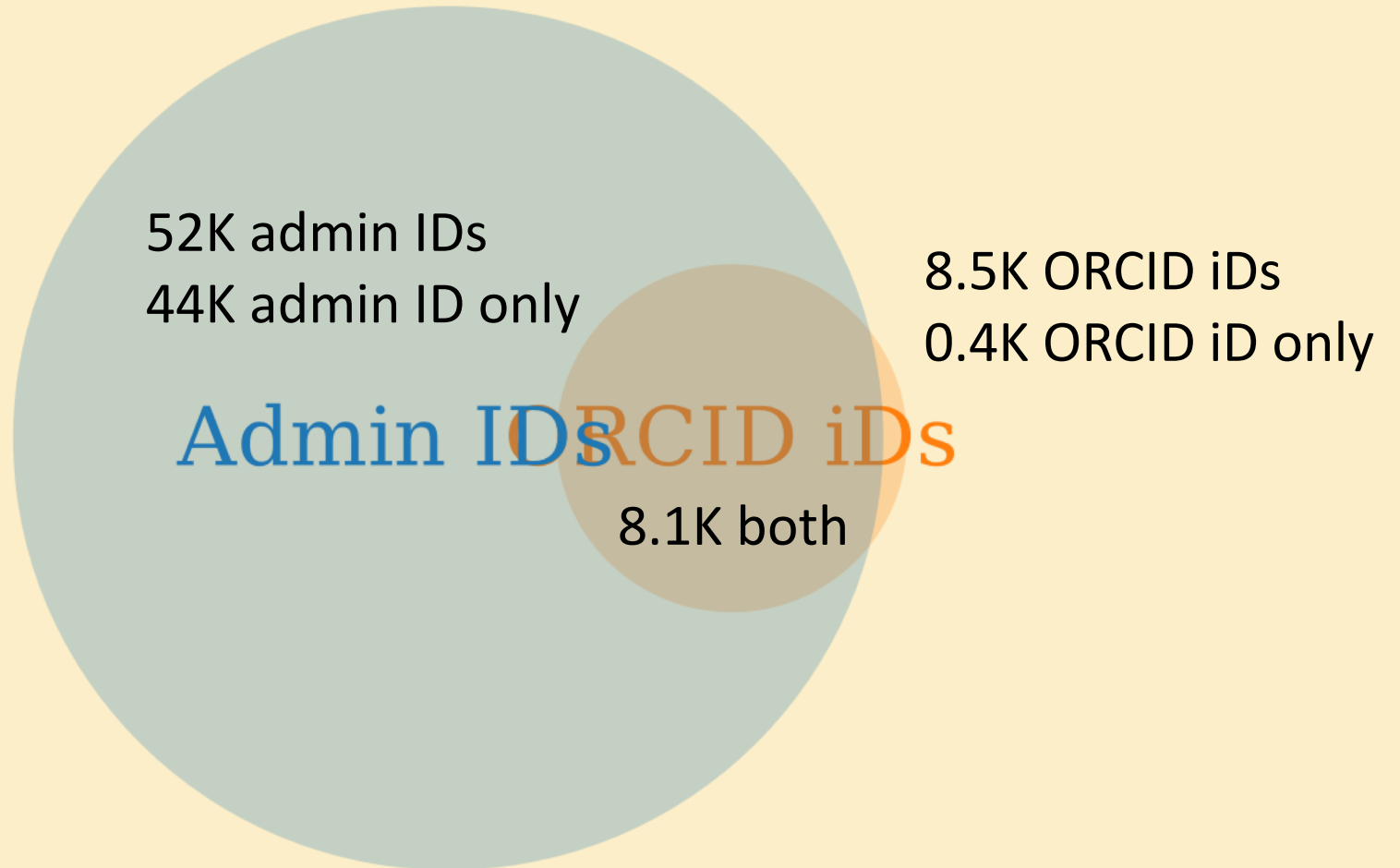
Babar Ali (státní příslušnost: PK - Pákistánská islámská republika, domácí tvůrce: A, orcid: [0000-0003-3811-214X](https://orcid.org/0000-0003-3811-214X))
Kamil Augsten (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 9957065, orcid: [0000-0001-8324-0576](https://orcid.org/0000-0001-8324-0576))
Petr Gallus (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 4553276)
Miroslav Havránek (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 1942158)
Zdeněk Hubáček (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 2141116, orcid: [0000-0003-3250-9066](https://orcid.org/0000-0003-3250-9066))
Miroslav Myška (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 4813197, orcid: [0000-0003-0982-3380](https://orcid.org/0000-0003-0982-3380))
Radek Novotný (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 3783467)
Petr Vokáč (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 6935192, orcid: [0000-0002-3429-4778](https://orcid.org/0000-0002-3429-4778))
Václav Vrba (státní příslušnost: CZ - Česká republika, domácí tvůrce: A, vedidk: 2190222, orcid: [0000-0002-0457-5675](https://orcid.org/0000-0002-0457-5675))
M. Aaboud (státní příslušnost: CZ - Česká republika)
G. Aad (státní příslušnost: CZ - Česká republika)
B. Abbott (státní příslušnost: CZ - Česká republika)
O. Abdinov (státní příslušnost: CZ - Česká republika)
Davide Caforio (státní příslušnost: CZ - Česká republika, orcid: [0000-0001-7808-8442](https://orcid.org/0000-0001-7808-8442))
Stanislav Pospíšil (státní příslušnost: CZ - Česká republika, orcid: [0000-0001-5424-9096](https://orcid.org/0000-0001-5424-9096))
Tomáš Slavíček (státní příslušnost: CZ - Česká republika, orcid: [0000-0001-8054-7345](https://orcid.org/0000-0001-8054-7345))
Karel Smolek (státní příslušnost: CZ - Česká republika, orcid: [0000-0002-5996-7000](https://orcid.org/0000-0002-5996-7000))
Michael Solar (státní příslušnost: CZ - Česká republika)
André Sopczak (státní příslušnost: CZ - Česká republika, orcid: [0000-0001-6981-0544](https://orcid.org/0000-0001-6981-0544))
Michal Suk (státní příslušnost: CZ - Česká republika)
Václav Vacek (státní příslušnost: CZ - Česká republika, orcid: [0000-0001-9584-0392](https://orcid.org/0000-0001-9584-0392))

2 years of publication data for the country



How many authors?

IS VaVal/RIV, collection years 2018–2019, Internals authors only



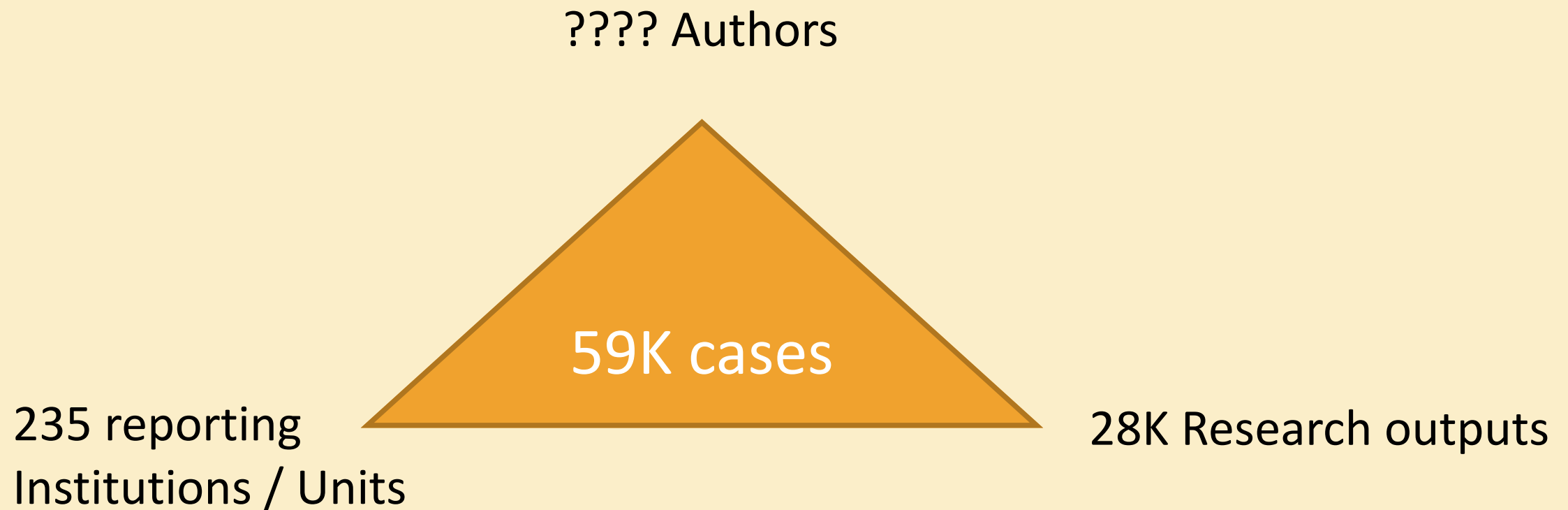
ORCID adoption rate in the country

- National CRIS: 8K
- ORCID knows of ~35K profiles registered with *.cz email address
- The ORCID Public Data Dump (October 2019) has ~14K profiles with CZ institution employment or education

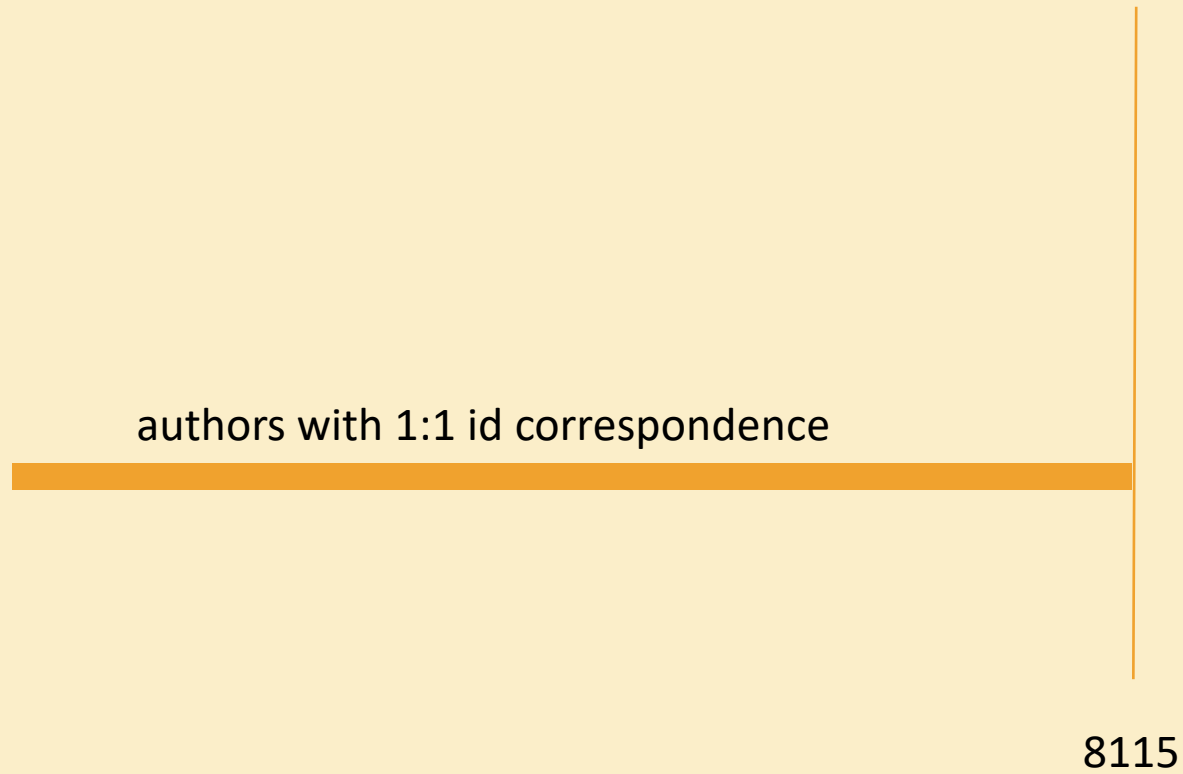
=> Est. 25–30K ORCID iDs, so ~1/2 researchers may have an ORCID

=> The national CRIS is having just ~1/4 of the ORCID iDs

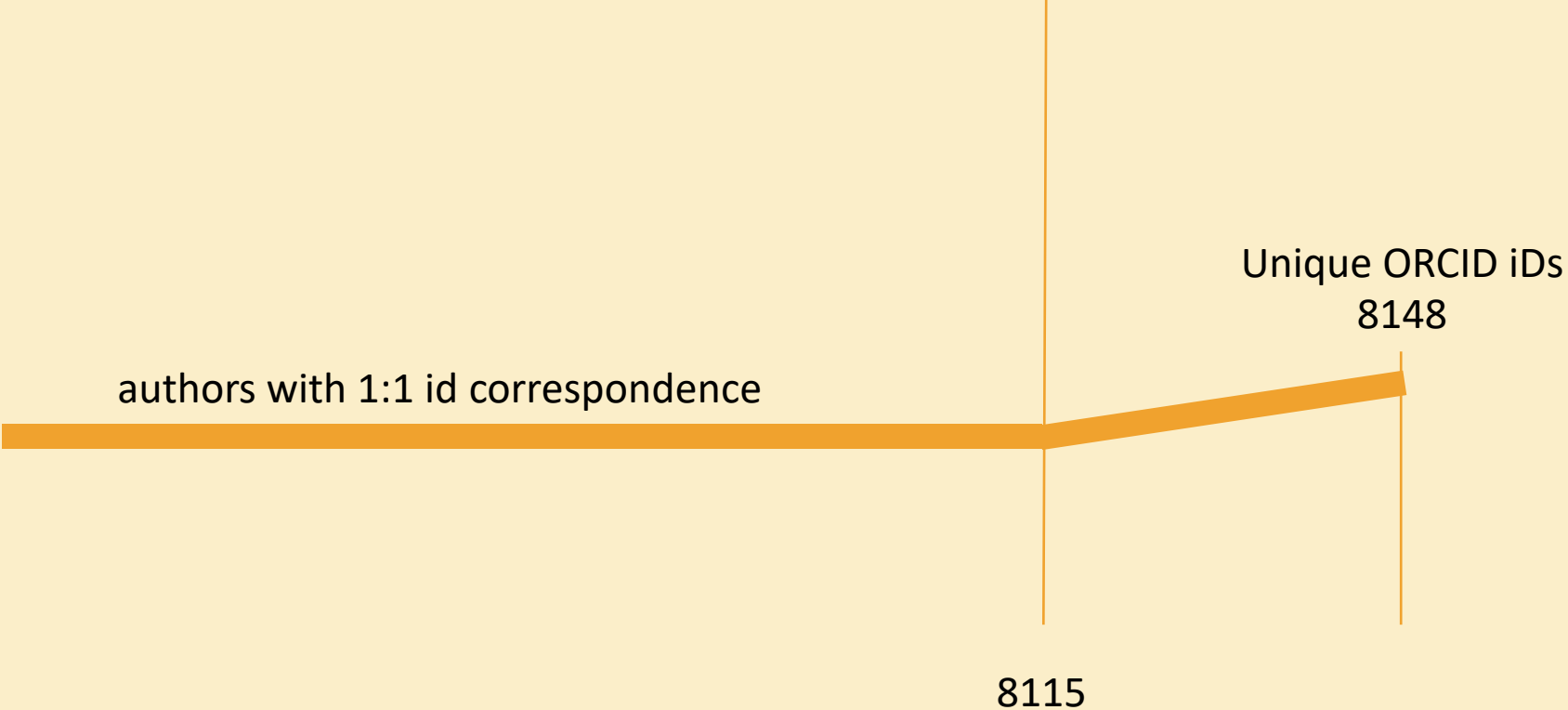
Authors with both identifiers



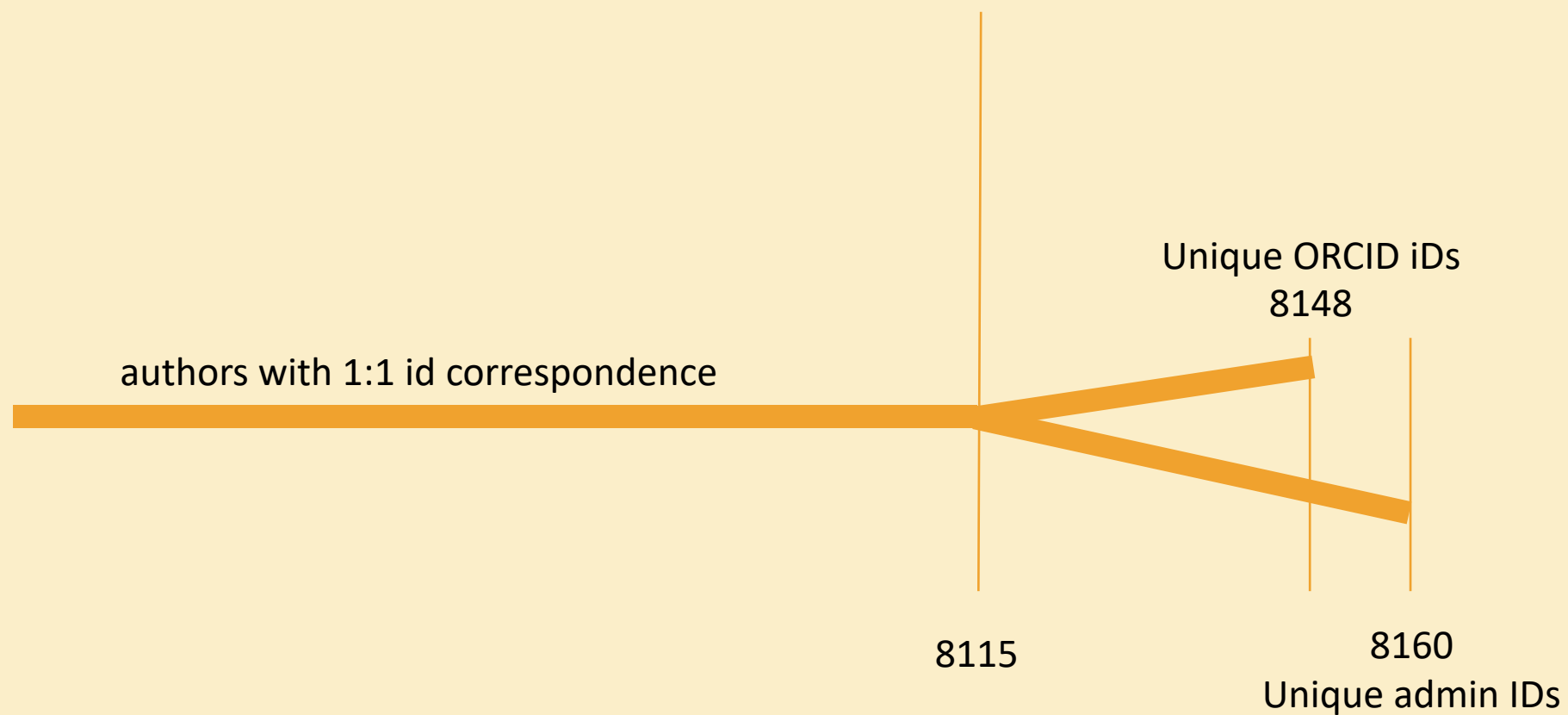
Authors with both identifiers



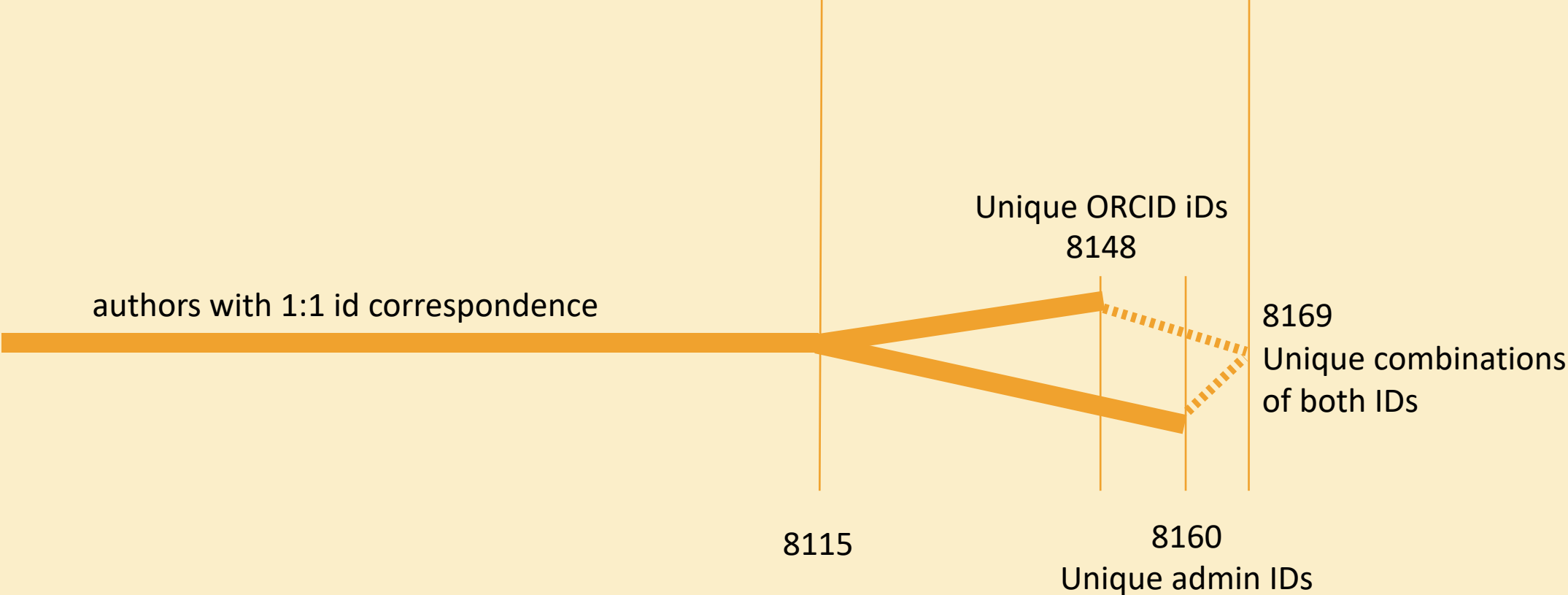
Authors with both identifiers



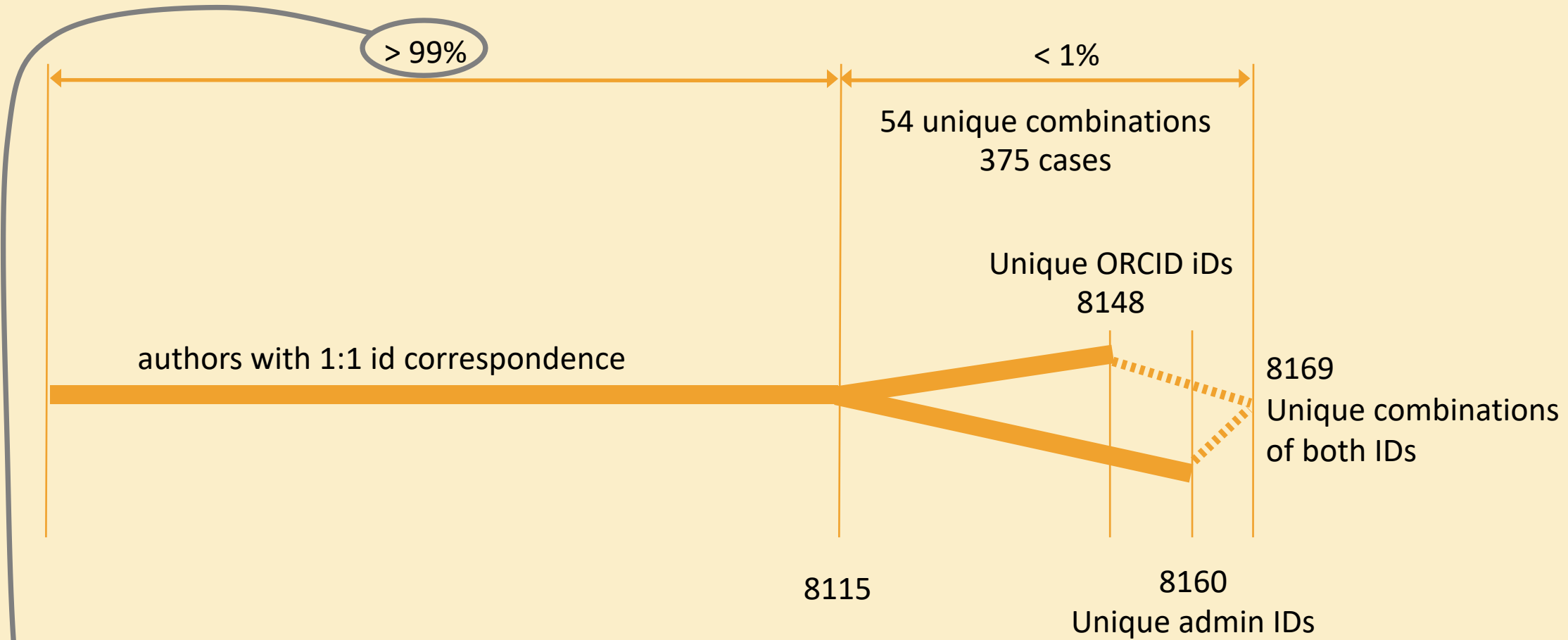
Authors with both identifiers



Authors with both identifiers



Authors with both identifiers



So a very good agreement overall!

Inspection of the non-1:1 cases surfaces:

1. False affiliation claims
(unsubstantiated by the affiliations listed on the publication)
2. The person having two different admin IDs with different reporting units
(should be impossible)
3. Wrong ORCID iD
(use of unauthenticated ORCID iDs entered by other staff, not the researcher; empty ORCID profiles particularly susceptible)

Inspection of the non-1:1 cases shows minimum:

1. Use of multiple ORCID iDs by the same researcher
(just 4 such researchers had a profile per affiliation)

Input from the PIDapalooza2020 audience:

Where can this be brought next?

- Get more ORCID iDs
 - *By increasing ORCID adoption (policy measures, mandated use)*
 - *By mining the ORCID Registry by the DOIs of earlier works*
- Talk to people in the Netherlands (Digital Author Identifiers)
 - *Experience (similar? different?)*
 - *Possible joint study*

Thank you!