AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

Piloting the European Unified Patient Identity Management (EUPID) Concept to Facilitate Secondary Use of Neuroblastoma Data from Clinical Trials and Biobanking

Ebner Hubert

Introduction

- Neuroblastoma most common cancer in infancy
- Rare disease – about 650 Incidents p.a. USA [1]
- Data collection during clinical trials needs to be done over large networks and many institutions

Secondary Use

- Re-Use of existing datasets from clinical trials
- Combine datasets from different sources

Secondary Use

- Re-Use of existing datasets from clinical trials
- Combine datasets from different sources

Aims

- Provide new findings
- Improve future treatment

Secondary Use

- Re-Use of existing datasets from clinical trials
- Combine datasets from different sources

Aims

- Provide new findings
- Improve future treatment

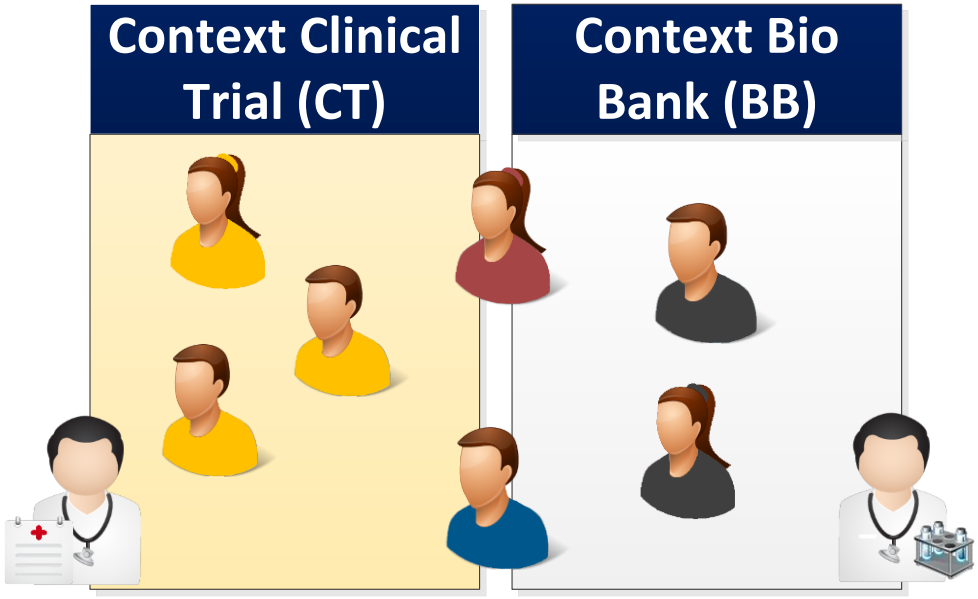
Hurdles

- Regulatory barriers
- Patient consent

Preparations for Secondary Use

- Pseudonymisation of Patients - Removing of patient identifiers from datasets
- Provide linkage patients over different trials
- Provide linkage of the accompanying datasets

Pilot Test Setting



European Unresectable Neuroblastoma (EUNB) trial

160 Patients

Results from array comparative genomic hybridisation (CGH) analyses from corresponding tumour samples

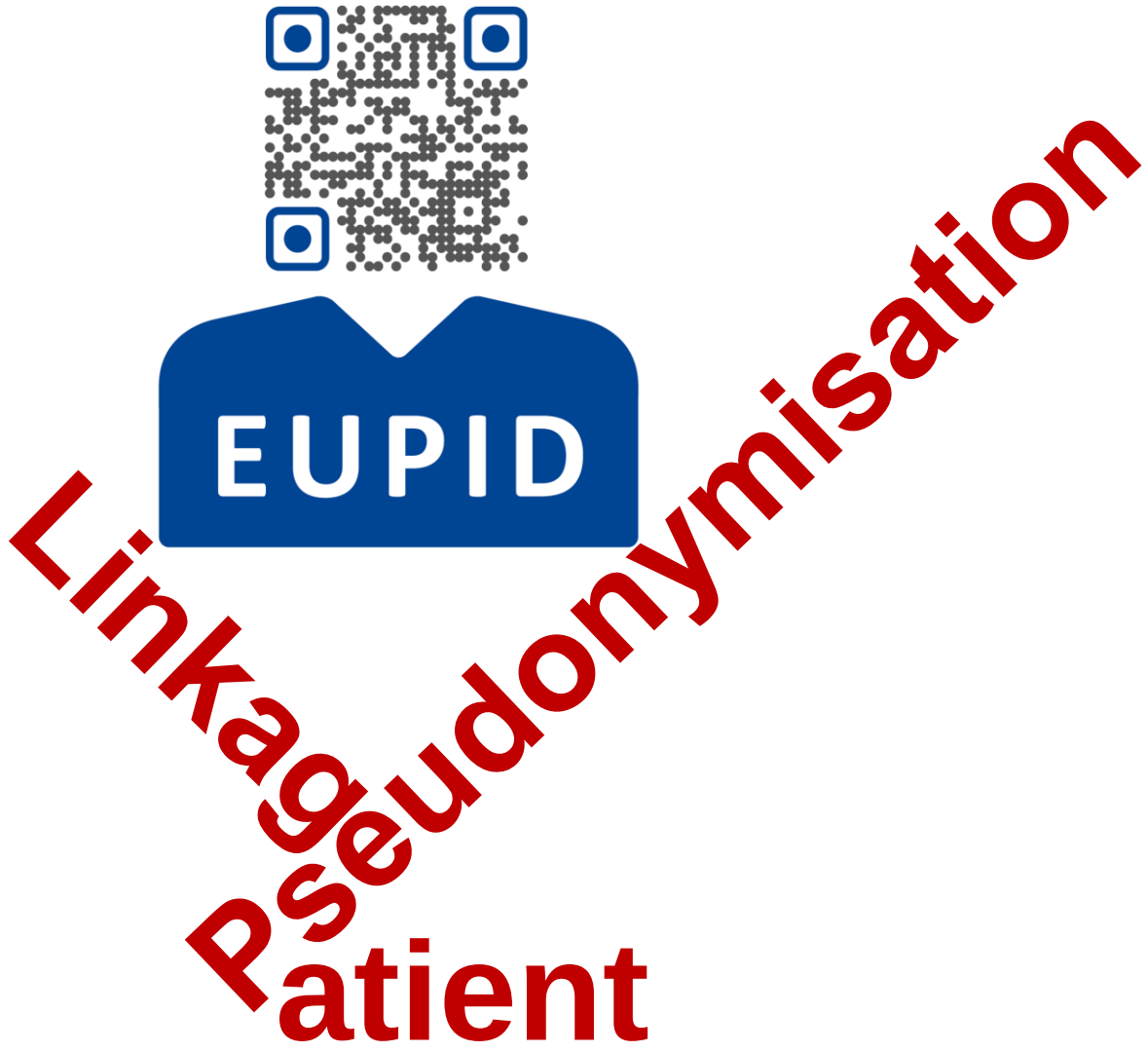
48 Patients

Needs for Secondary Use in the Pilot Test

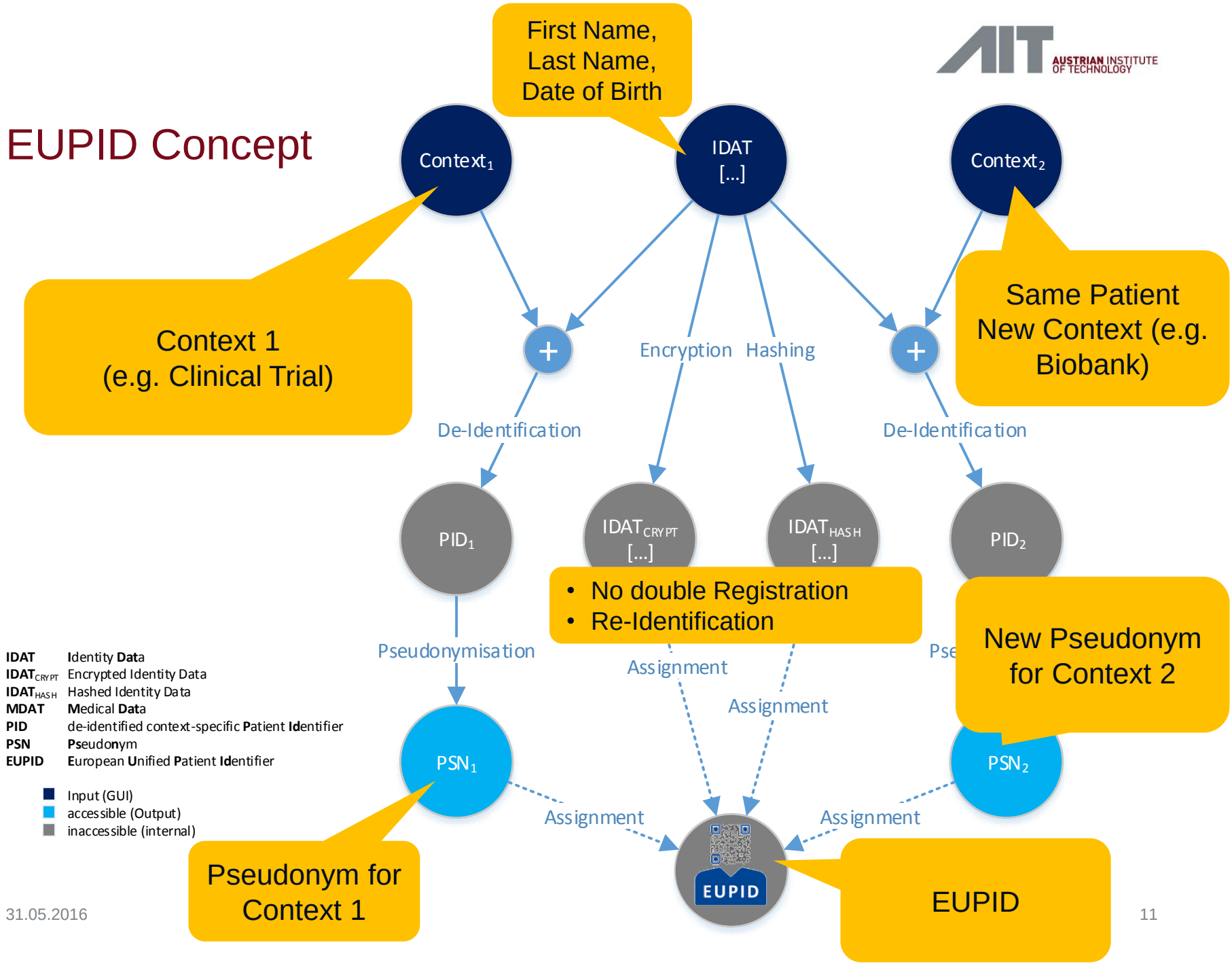
- Creation of Pseudonyms
  - Pseudonymisation of patients in Clinical Trial (CT)
  - Pseudonymisation of patients in Bio Banking (BB)

- Linkage of patients
- Linkage of accompanying datasets

European Unified Patient Identity Management



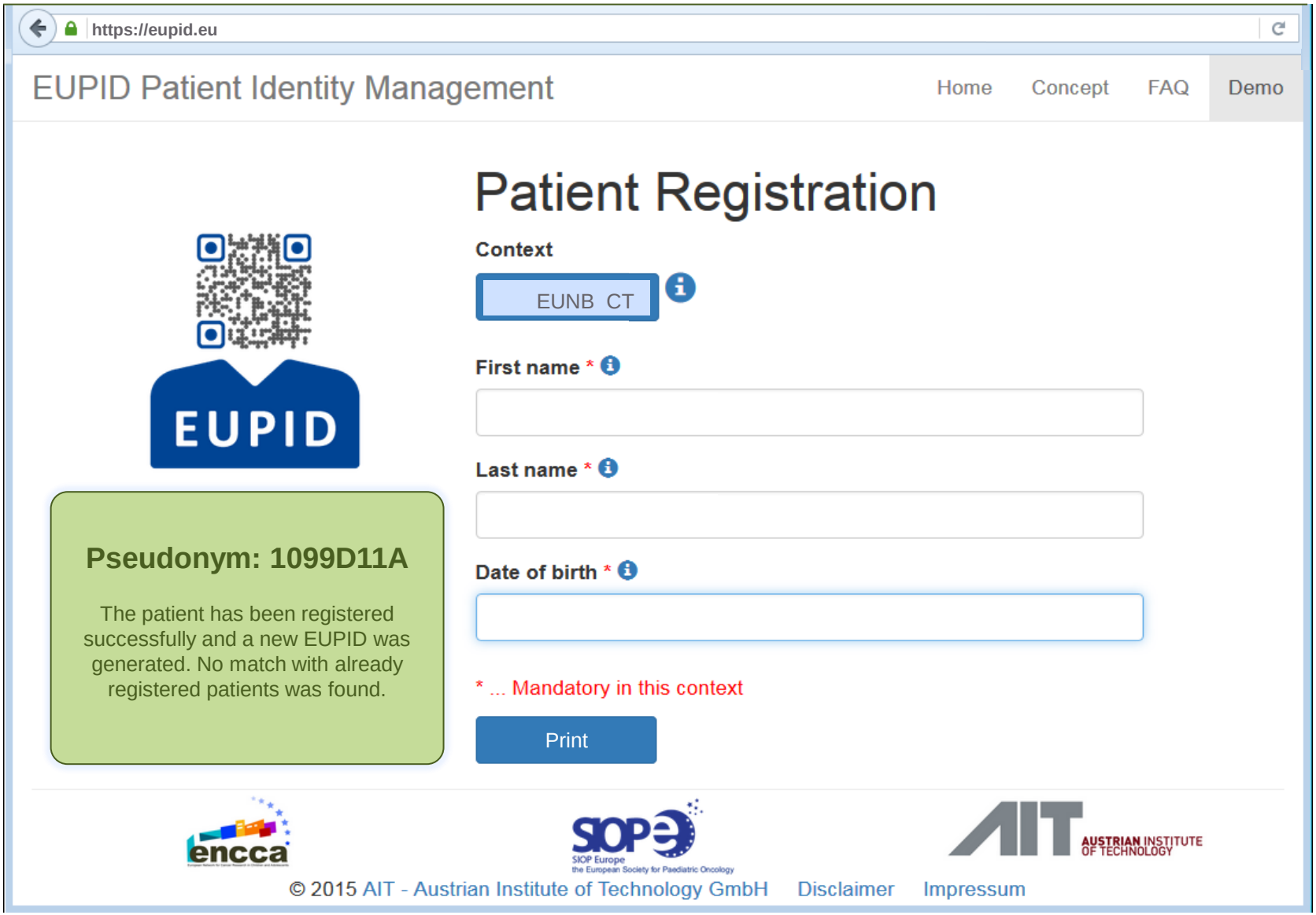
EUPID Concept



IDAT	**I**dentity **Dat**a
IDAT$_{CRYPT}$	Encrypted Identity Data
IDAT$_{HASH}$	Hashed Identity Data
MDAT	**M**edical **Dat**a
PID	de-identified context-specific **P**atient **Id**entifier
PSN	**Ps**eudo**n**ym
EUPID	**E**uropean **U**nified **P**atient **Id**entifier

Input (GUI)
accessible (Output)
inaccessible (internal)

EUPID - Client

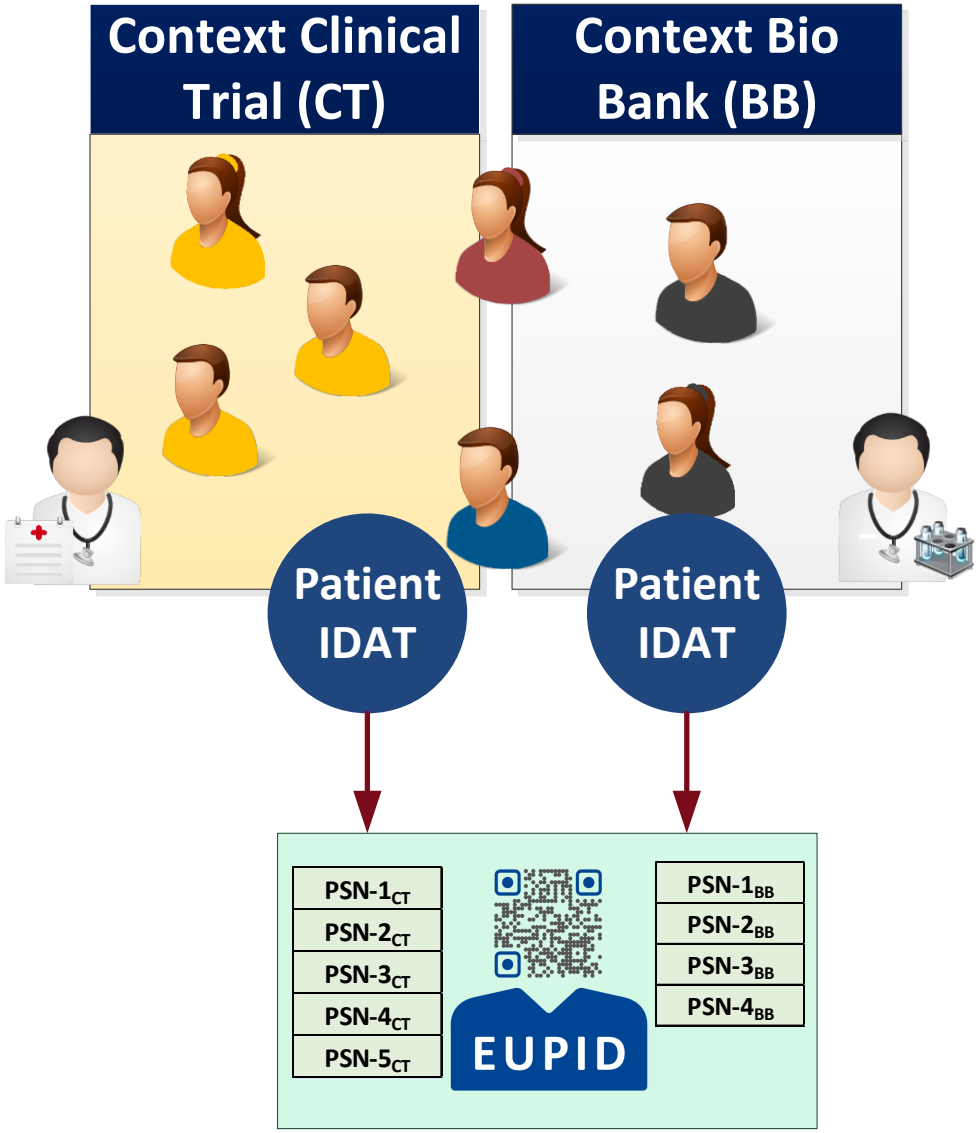


EUPID - Client

- Webapp on local client
- Plain text IDAT never leaves client
- Encrypting and phonetic hashing of IDAT done locally

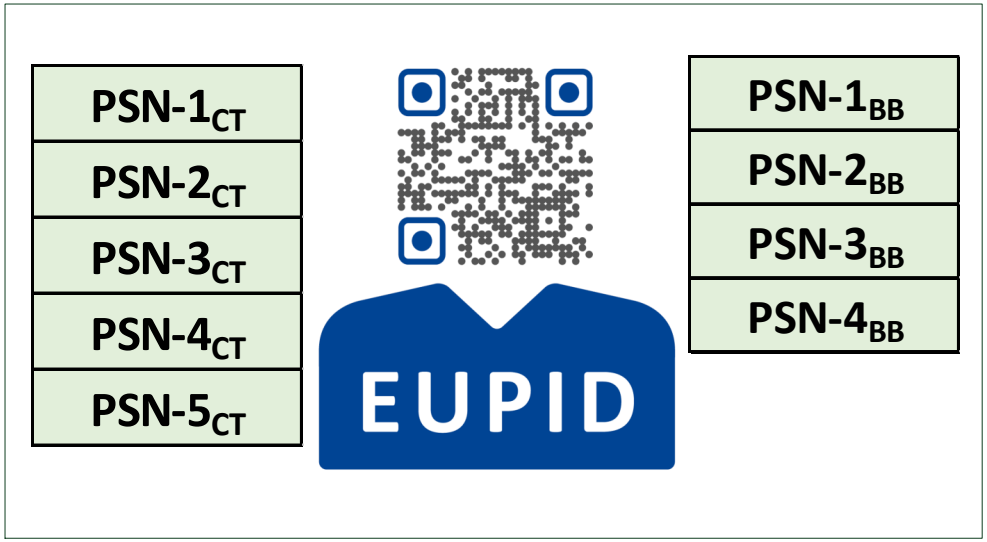
- Only Encrypted and Hashed Values are transferred

EUPID - Pseudonymisation



IDAT…Identification Data
PSN… Pseudonyms

EUPID – Linkage



Reference Table	
PSN_{CT}	PSN_{BB}
$PSN\text{-}1_{CT}$	$PSN\text{-}2_{BB}$
$PSN\text{-}2_{CT}$	
$PSN\text{-}3_{CT}$	$PSN\text{-}3_{BB}$
$PSN\text{-}4_{CT}$	

PSN… Pseudonyms

Needs for Secondary Use in the Pilot

- Creation of Pseudonyms
  - Pseudonymisation of patients in Clinical Trial (CT) ✓
  - Pseudonymisation of patients in Bio Banking (BB) ✓
- Linkage of Patients (if there is a link) ✓
- Linkage of accompanying data

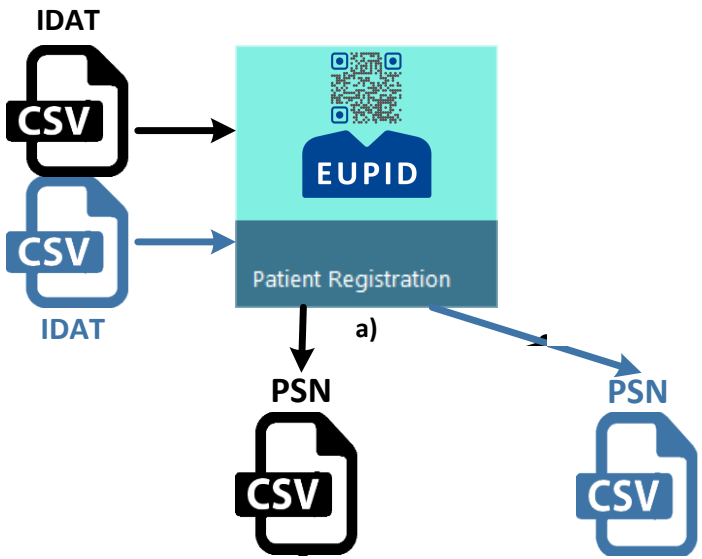
ABCD-4-E pipeline for data production

- EUPID Patient Registration
- Standardization of Data
  (Operational Data Model [14], XML-based)
- PSN-Exchange
- Data upload to Repository

ABCD-4-E … Advanced Biomedical Collaboration Domain for ENCCA [4]
ENCCA … European Network for Cancer Research in Children and Adolescents [5]

EUPID - Patient Registration

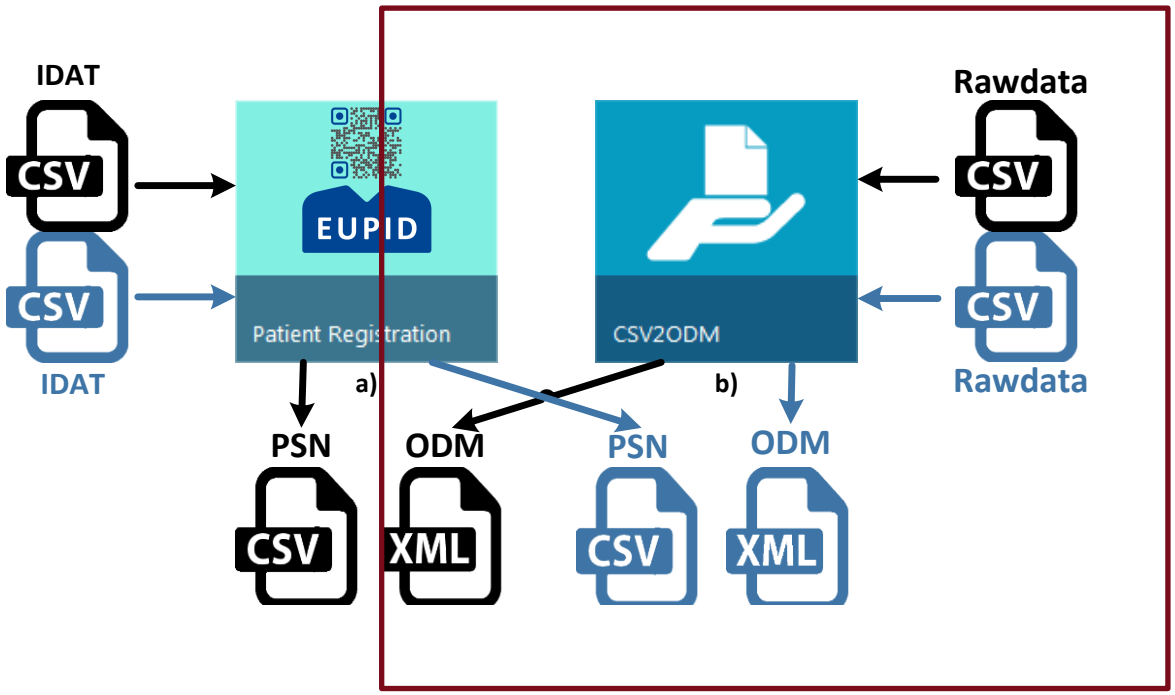
- **CLINICAL TRIAL (160 patients)**
- **BIOBANK (48 patients)**



IDAT … Identification Data
PSN … Pseudonym

CSV2ODM – Standardization of Data

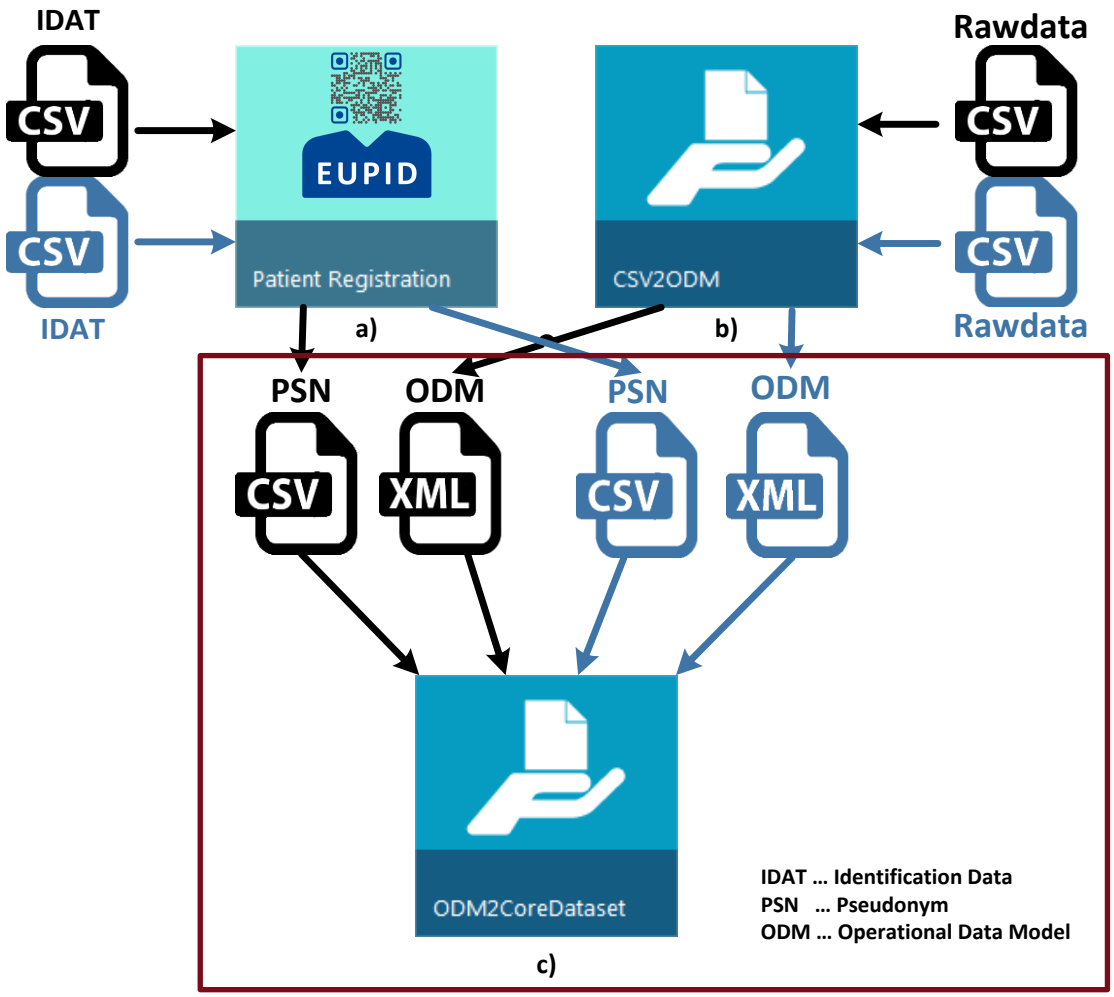
- **CLINICAL TRIAL (160 patients)**
- **BIOBANK (48 patients)**



IDAT … Identification Data
PSN … Pseudonym
ODM … Operational Data Model

ODM2CoreDataset - PSN-Exchange and Data upload

- **CLINICAL TRIAL (160 patients)**
- **BIOBANK (48 patients)**



Results

- EUPID was able to link each of the 48 patients from the biobank (BB) to their corresponding pseudonym from the clinical trial (CT).
- Joined dataset containing linked pseudonyms for 48 patients
- AIT took over the role of the trusted third party to access the *Reference Table*.

Discussion

- EUPID overcomes barriers of regulatory framework and patient consent
- Re-identification possible by Trusted Third Party
- Strategies to avoid Re-identification by comparing datasets
  - Randomizing order of datasets rows
  - Coarsening of typical individual information
  - k-anonymity
  - l-diversity

References

[1] P. P. T. E. Board., "Neuroblastoma Treatment (PDQ®): Health Professional Version.," in PDQ Cancer Information Summaries, ed: National Cancer Institute, 2016.

[2] (2016-01-19). SIPEN-R-NET. Available: http://www.siopen.org

[3] BMA. (2014, 2016-01-19). requests for disclosure of data for secondary purposes. Available: http://www.bma.org.uk/-/media/files/pdfs/practical%20advice%20at%20work/ethics/releasingdataforsecondaryuses.pdf

[4] (2016-01-19). ENCCA - European Network for Cancer Research in Children and Adolescents. Available: www.encca.eu

[5] M. Nitzlnader, M. Falgenhauer, C. Gossy, and G. Schreier, "Architecture for an advanced biomedical collaboration domain for the European paediatric cancer research," Stud Health Technol Inform, pp. 167-74, 2015.

[6] M. Nitzlnader and G. Schreier, "Patient identity management for secondary use of biomedical research data in a distributed computing environment," Stud Health Technol Inform, pp. 211-8, 2014.

[7] J. A. Kohler, H. Rubie, V. Castel, K. Beiske, K. Holmes, C. Gambini, et al., "Treatment of children over the age of one year with unresectable localised neuroblastoma without MYCN amplification: Results of the SIOPEN study," European Journal of Cancer, vol. 49, pp. 3671-3679, Nov 2013.

[8] G. Schleiermacher, J. Michon, A. Ribeiro, G. Pierron, V. Mosseri, H. Rubie, et al., "Segmental chromosomal alterations lead to a higher risk of relapse in infants with MYCN-non-amplified localised unresectable/disseminated neuroblastoma (a SIOPEN collaborative study)," British Journal of Cancer, vol. 105, pp. 1940-1948, Dec 6 2011.

[9] NCBI. (2015, 12.01.2016). Gene Expression Omnibus. Available: http://www.ncbi.nlm.nih.gov/geo/

[10] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," Nucleic Acids Research, vol. 30, pp. 207-210, Jan 1 2002.

[11] (2016-01-19). array-CGH profiling of human neuroblastoma samples obtained from infants included in the INES99.1, INES99.2 and INES99.3 trials. Available: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26494

[12] (2016-01-19). The International Neuroblastoma Risk Group (INRG). Available: http://inrgdb.org

[13] S. L. Cohn, A. D. J. Pearson, W. B. London, T. Monclair, P. F. Ambros, G. M. Brodeur, et al., "The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report," Journal of Clinical Oncology, vol. 27, pp. 289-297, Jan 10 2009.

[14] C. D. I. S. Consortium. (2016-01-19). Operational Data Model. Available: http://www.cdisc.org/odm

[15] N. Hochedlinger, D. Nitzlnader, M. Falgenhauer, S. Welte, D. Hayn, L. Koumakis, et al., "Standardized data sharing in a paediatric oncology research network--a proof-of-concept study," Stud Health Technol Inform, pp. 27-34, 2015.

[16] A.-A. I. o. T. GmbH. (2015-01-19). European Patient Identity Management EUPID. Available: https://eupid.eu

[17] G. Danezis, J. Domingo-Ferrer, M. Hansen, J.-H. Hoepman, D. Le Métayer, R. Tirtea, et al., "Privacy and Data Protection by Design – from policy to engineering," ed: European Union Agency for Network and Information Security (ENISA), 2014

[18] L. Sweeney, "K-anonymity: A model for protecting privacy," International Journal on Uncertainty, fuzziness and knowledge-based systems, vol. 10, pp. 557 - 570, 2002.

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

Thank you

Ebner Hubert
hubert.ebner.fl@ait.ac.at