



CANDY

Comorbid Analysis of Neurodevelopmental Disorders and epilepsy

Deliverable D5.5

Report on status of posting results MPX

Grant Agreement number: 847818

SC1-BHC-01-2019

Start date of project: 1 January 2020

Duration: 66 months

Lead beneficiary:

IP

Due date of deliverable: 30/06/2025

Actual submission date: 28/08/2025

Document type: Report

Dissemination Level

PU Public

Table of Contents

1	Executive summary	3
2	Introduction.....	3
3	Main section.....	4
3.1	The CANDY-PIP and CANDY-MPX data in the OWEY datalake.....	4
3.1.1	CANDY-PIP Data	4
3.1.2	CANDY-MPX Data.....	5
3.1.1	Detail of Clinical questionnaires used in CANDY-PIP and CANDY-MPX studies	5
3.2	FAIR principles compliance of the data sharing of the CANDY-PIP and CANDY-MPX data .	8
3.2.1	Findable Data implementation	8
3.2.2	Accessible Data Framework	9
3.2.3	Interoperable Data Structure	9
3.2.4	Reusable Data Standards	9
4	Conclusions.....	9
5	Glossary.....	10
6	Bibliography.....	11

1 Executive summary

This deliverable **D5.5 Report on status of posting results MPX** present how the dataset has been collected, stored and will be shared with the scientific community. It also details the data collected. We present the OWEY datalake of the Institut Pasteur and how we used the FAIR principles to format and homogenize the dataset.

2 Introduction

This multi-centric and highly multimodal project required a specialized platform capable of securely handling large datasets in a GDPR-compliant manner, and across several European research institutions. To this end, the OWEY data lake, physically hosted at Institut Pasteur, was deployed (Figure 1). OWEY automates critical operations preceding data sharing: **Automated Participant Sharing Consent Verification, Double Pseudonymization, Automated File Organization and Traceability as well as Fine-Grained Access Management**. Finally, OWEY offers **Horizontal Scalability** since the system supports increasing volumes of data, users, and concurrent connections through a distributed and flexible architecture. These processes are integrated into a seamless system that ensures regulatory compliance, interoperability, and operational efficiency.

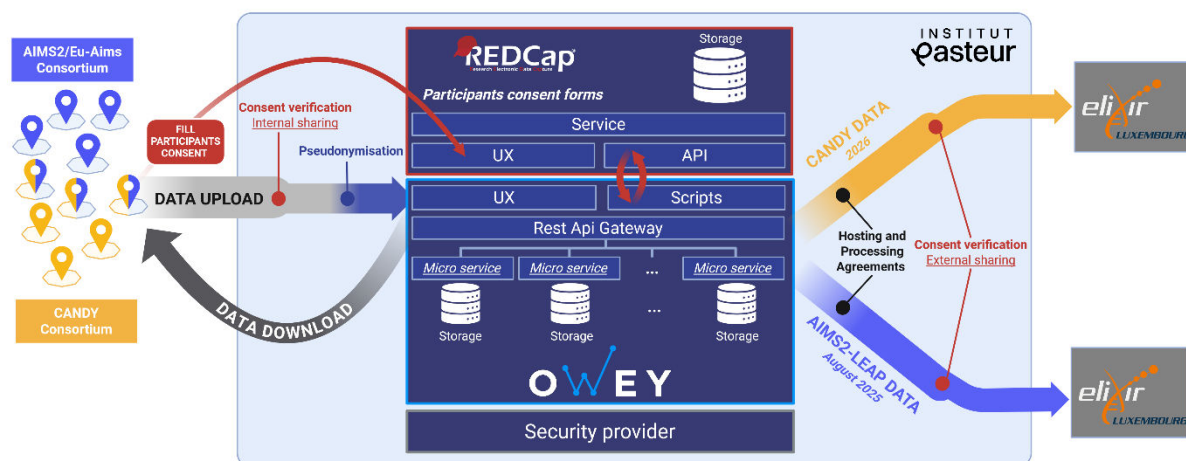


Figure 1: Schematic representation of the data flows and data storage infrastructure for Aims2-Trials LEAP study (top, blue) and CANDY study data (bottom, orange). Each participant's consent must be collected by each center concomitantly with their data. This consent data is stored in a dedicated REDCap (Harris et al., 2009) form hosted at Institut Pasteur. During the data upload from one center, data is automatically filtered based on participants consent for internal sharing (i.e. within the Eu-aims or CANDY consortium), verified using the REDCap API. The participants' identifiers are also pseudonymized, i.e. get attributed to a unique and anonymous identifier in OWEY. These steps allow each center to securely access and download data from all other centers and across modalities during the entire study run. For external sharing of LEAP data, it required each participating research center to sign a data hosting and processing agreement with ELIXIR-LU, a European life sciences infrastructure offering hosting services. Prior to data transfer from OWEY to ELIXIR-LU, data was filtered a second time based on the participants consent to share their data externally (i.e. outside the Eu-aims consortium). A similar procedure will be followed for the external sharing of the CANDY consortium data in 2026.

3 Main section

3.1 The CANDY-PIP and CANDY-MPX data in the OWEY datalake

In this section we provide a snapshot of the current data stored within the OWEY datalake. It includes the [CANDY-PIP](#) (Table 1) and the [CANDY-MPX](#) (Table 2) data sets. We also provide details on the clinical data collected (Table 3).

3.1.1 CANDY-PIP Data

Table 1: Summary of the CANDY-PIP data uploaded to the OWEY datalake and shared with the consortium members.

CANDY-PIP study	Ghent	London	Nijmegen	Paris	Stockholm	TOTAL
Clinical	28	37	38	28	22	153
Wave 1	9	15	10	8	4	46
Wave 2	19	22	28	20	18	107
BOLD	8	3	8	3	7	29
Wave 1	3	1	3	/	1	8
Wave 2	5	2	5	3	6	21
Anatomy	16	8	9	13	9	55
Wave 1	5	2	4	4	2	17
Wave 2	11	6	5	9	7	38
DTI	14	1	5	7	8	35
Wave 1	5	1	3	2	1	12
Wave 2	9	/	2	5	7	23
EEG	15	20	23	10	10	78
Wave 1	5	9	9	/	2	25
Wave 2	10	11	14	10	8	53
MRS	14	8	8	8	8	46
Wave 1	4	2	3	3	1	13
Wave 2	10	6	5	5	7	33
Eye-tracking	23	19	25	15	16	98
Wave 1	8	4	10	4	3	29
Wave 2	15	15	15	11	13	69
Touchscreen	25	18	24	14	16	97
Wave 1	9	4	8	6	3	30
Wave 2	16	14	16	8	13	67
Inflammatory proteomic						26
Immune Brain Gut (Wave 2)						16
Inflammation (Wave 2)						10

WGS total	48	28	31	21	38	166
<i>Proband</i>	19	9	12	8	12	60
<i>Parents</i>	29	19	19	13	26	106

3.1.2 CANDY-MPX Data

Silos dedicated to the CANDY-MPX study were created on the OWEY datalake. Details of the data uploaded are visible in the table 2.

Table 2: Summary of the CANDY-MPX data uploaded to the OWEY datalake and shared with the consortium members.

CANDY-MPX study	Ghent	London	Nijmegen	Paris	Stockholm	TOTAL	SPX individuals - Créteil
Clinical	47	58	47	342	31	525	44
<i>Proband</i>	10	12	10	65	10	107	
<i>Relatives</i>	37	46	37	277	21	418	
WGS	46	29	36	341	34	486	41
Inflammatory proteomic						99	
<i>Immune Brain Gut</i>						69	
<i>Inflammation</i>						30	

3.1.1 Detail of Clinical questionnaires used in CANDY-PIP and CANDY-MPX studies

Table 3: Description and references for clinical questionnaires in alphabetical order. W1 and W2 indicates if the data was collected in Wave 1 and W2 of the PIP study, respectively. Data was collected by clinicians except if self-report is specified. In that case, surveys were sent to the participants (parents or caregivers).

Questionnaire Name	CANDY-PIP	CANDY-MPX	Description	Reference
ADHDRS (ADHD Rating Scale)		Self-report, children and parents	Assesses symptoms of attention-deficit/hyperactivity disorder (ADHD) in children or adults.	(DuPaul et al., 1998)
ADOS	W1, W2		Autism Diagnostic Observation Schedule (ADOS) for assessing autism spectrum disorder (ASD)	(Lord et al., 2012)
<i>ADOS Module 1</i>			For children with limited verbal skills.	
<i>ADOS Module 2</i>			For children with phrase speech but limited conversational abilities.	
<i>ADOS Module 3</i>			For verbally fluent children.	

Attention and Activity Levels	W1, W2 (Self-report)		Evaluates attention span, hyperactivity, and impulsivity for screening preschoolers with behavior problems.'	(McGoey et al., 2007)
Autism Diagnostic Interview Revised (ADI-R)	W1, W2		A structured interview for diagnosing ASD and gathering detailed developmental history.	(Lord et al., 1994)
Autistic Disorder Assessment			Evaluates symptoms of autism spectrum disorder (ASD) according to the DSM-5 criterias.	(American Psychiatric Association, 2022)
Child Routines Inventory Revised	W1, W2 (Self-report)		Evaluates daily routines and habits in children.	(Sytsma et al., 2001)
Children's Behaviour Questionnaire (CBQ)	W1, W2 (Self-report)		Parent-report instrument, allowing the highly differentiated assessment of temperament in early to middle childhood	(Rothbart et al., 2001)
Child's Sleep Habits Questionnaire (CSHQ)	W1, W2 (Self-report)	Self-report	Assesses sleep patterns, disturbances, and quality in children.	(Owens et al., 2000)
Clinical Global Impression (CGI)	W2		Clinician-rated scale for assessing overall severity of condition, compared to the previous visit of last 12-15 months.	
COVID-19 Questionnaire	W1 (all), W2 (ADHD only) (Self-report)		Collects information on COVID-19 exposure, infection, and impact on the child and family life.	
Demographic and Family questionnaire	W1, W2		Concise demographic information about the participant (e.g., age, gender, ethnicity) and family structure, socioeconomic status, and background.	
Diet Questionnaire	W1, W2 (Self-report)		Records dietary habits, and nutritional intake of the child.	
Epilepsy and Seizure Questionnaire	W1, W2 (Self-report)		Evaluates history of epilepsy, seizure frequency, type, and impact on the child's life.	
Family Questionnaire / Family History Interview (FHI)	W2		Gathers information about family composition and genealogy, presence of neurodevelopmental disorders, common mental health conditions, chromosomal or genetic disorders.	
Febrile Seizure Questionnaire	W1, W2 (Self-report)		Records history of febrile seizures of the child.	
Global Assessment of Functioning (GAF)			Clinician-rated scale for assessing overall psychological, social, and occupational functioning.	(Hall, 1995)

Health and Life Events	W1, W2 (Self-report)		Records significant health events and life changes in the pregnancy, and during the child's first year of life, and during the last 6 months.	
Highly Sensitivity Child Scale	W1, W2 (Self-report)		Self-report measure of Environmental Sensitivity in children and adolescents.	(Pluess et al., 2018)
HRS-MAT (Caregiver)	W2		Adaptive measure of nonverbal IQ, with psychometric properties comparable to the Raven's Standard Progressive Matrices (RSPM).	(Hansen, 2019)
Intellectual Efficiency Adaptive Behavior			Assesses cognitive functioning and adaptive skills : IQ test(s) and Vineland (see below).	
Intervention History and Medication	W1, W2 (Self-report)		Documents past and current interventions or therapies the child has received. Details about medication received, if applicable (name, dosage, start and end date of treatment).	
Intolerance Of Uncertainty	W1, W2 (Self-report)		Assesses anxiety related to uncertainty and unpredictability.	(Boulter et al., 2014)
KINDL	W2 (Self-report)		The KINDL ^R is a generic instrument for assessing Health-Related Quality of Life in children and adolescents aged 3 years and older	(Ravens-Sieberger & Bullinger, 1998)
Medical History	W1, W2 (Self-report)	Self-report	Records the child's medical history, including pregnancy and childbirth details, firsts acquisitions, any other medical conditions and epilepsy presence.	
Mullen	W1, W2		Mullen Scales of Early Learning assesses cognitive and motor development in young children.	(Mullen, 1995)
NART (Caregiver)	W2		NART for assessing cognitive ability in the caregiver(s).	(Nelson et al., 2022)
Physical Examination			Records clinical observations (e.g., height, weight, lateralization, vital signs) during a medical assessment.	
Pre Perinatal	W1 (all), W2 (ADHD only) (Self-report)		Collects information about prenatal and perinatal factors (e.g., health during pregnancy, labour and delivery, including complications, and weight at birth).	
PSQI (Pittsburgh Sleep Quality Index)		Adult relatives (Self-report)	Evaluates sleep quality and disturbances in adults.	(Buysse et al., 1989)

Raven's Progressive Matrices (Raven)		Self-report	Non-verbal test of fluid intelligence and problem-solving ability.	(Raven, 1938)
Sensory Experiences Questionnaire	W1, W2 (Self-report)		Measures sensory processing differences in children.	(Baranek et al., 2006)
Short Sensory Profile (SSP)		Self-report	Measures sensory processing differences in children (abbreviated version of the Sensory Profile).	(Tomchek & Dunn, 2007)
Social Responsiveness Scale (SRS)	W1, W2		Measures social communication and interaction skills in children and adults. P	(Constantino & Gruber, 2012)
Strengths And Difficulties P24	W1, W2		A screening tool to differentiate low, medium, and high levels of emotional or behavioral difficulties (parent-report).	(Bourdon et al., 2005)
Vineland 3	W1, W2		Assesses adaptive behavior skills (communication, daily living, socialization, motor skills) in children.	(Sparrow et al., 2016)

3.2 FAIR principles compliance of the data sharing of the CANDY-PIP and CANDY-MPX data

The data management strategy for the CANDY research project encompasses the collection and storage of harmonized datasets, secure hosting and sharing with research partners, and post-publication dissemination to the scientific community (external data sharing). Our strategy adheres to the **FAIR principles** - ensuring data are **Findable, Accessible, Interoperable, and Reusable** (Wilkinson et al., 2016), reflecting best practices established through our previous LEAP study within the AIMS2 trials consortium (Loth et al., 2017).

3.2.1 Findable Data implementation

Rich metadata describing study parameters, participant characteristics, and data modalities are systematically captured and indexed within the OWEY data lake architecture, enabling efficient search and retrieval across the multi-centric, multimodal dataset. This metadata will be shared along with the data externally (outside the consortium) in ELIXIR-LU servers, as recently done for the LEAP study in the AIMS2-Trials consortium (Figure 1).

Data discoverability will then be achieved through the generation of globally unique and persistent identifiers assigned to all datasets after external sharing (DOI).

3.2.2 Accessible Data Framework

The OWEY platform ensure data retrievability while maintaining security compliance for exchanges within the consortium. It includes **automated participant sharing consent verification**, dynamically controlling data access based on current consent status. Access to data is contingent on the presence of valid consent, collected using the REDCap electronic data capture tools hosted at the Institut Pasteur (Harris et al., 2009). If a participant withdrawn his consent to share data inside the research consortium, the corresponding data is hidden from the users within OWEY, ensuring dynamic regulatory compliance.

Our **GDPR-compliant infrastructure**, hosted at Institut Pasteur, also provides authentication and authorization procedures that allow authorized consortium members to access datasets through secure, standardized interfaces. Similarly, the ELIXIR-LU data portal provides **robust, compliant access pathways** for the broader scientific community.

3.2.3 Interoperable Data Structure

In our OWEY datalake, the **automated file organization system structures data according to multi-axis schemes** (study, site, modality, analysis level), enabling integration across different analytical platforms and research environments. This systematic approach supports the combination of diverse data modalities - clinical, cognitive, neuroimaging, and genetic data - facilitating cross-platform analysis.

3.2.4 Reusable Data Standards

Data reusability is maximized through comprehensive documentation and clear usage licensing. Each dataset shared externally (i.e. using ELIXIR-LU) will include detailed provenance information, quality metrics, and contextual metadata that enable accurate interpretation and replication in new research contexts. The ELIXIR-LU dataset portal provides transparent access mechanisms with explicit licensing terms, supporting restricted access paradigms adapted to our data sensitivity level.

4 Conclusions

Altogether, the established framework from the LEAP study from AIMS2-Trials, as presented in Figure 1, has been directly applied to the CANDY project, ensuring **consistent FAIR compliance** across our research portfolio and facilitating future data sharing initiatives beyond the current consortium. The multiscale datasets collected during the CANDY project are securely stored in OWEY, the Data Lake of the Institut Pasteur and ready to be shared to the scientific community.

5 Glossary

Abbreviation / acronym	Description
MPX	Multiplex
W	Wave

6 Bibliography

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.).

Baranek, G. T., David, F. J., Poe, M. D., Stone, W. L., & Watson, L. R. (2006). Sensory Experiences Questionnaire: Discriminating sensory features in young children with autism, developmental delays, and typical development. *Journal of Child Psychology and Psychiatry*, 47, 591–601.

Boulter, C., Freeston, M., South, M., & Rodgers, J. (2014). Intolerance of uncertainty as a framework for understanding anxiety in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 44(6), 1391–1402.

Bourdon, K. H., Goodman, R., Rae, D. S., Simpson, G., & Koretz, D. S. (2005). The Strengths and Difficulties Questionnaire: U.S. Normative Data and Psychometric Properties. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Buyse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research*, 28(2), 193–213.

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale (SRS)*. Western Psychological Services.

DuPaul, G. J., Power, T. J., Anastopoulos, A. D., & Reid, R. (1998). *ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation*. Guilford Press.

Hall, R. C. (1995). Global assessment of functioning: A modified scale. *Psychosomatics*, 36(3), 267–275.

Hansen, J. A. (2019). Development and psychometric evaluation of the Hansen Research Services Matrix Adaptive Test: A measure of nonverbal IQ. *Journal of Autism and Developmental Disorders*, 49(7), 2721–2732.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, 42(2), 377–381. <https://doi.org/10.1016/j.jbi.2008.08.010>

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2)*. Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*.

Loth, E., Charman, T., Mason, L., Tillmann, J., Jones, E. J. H., Wooldridge, C., Ahmad, J., Auyeung, B., Brogna, C., Ambrosino, S., Banaschewski, T., Baron-Cohen, S., Baumeister, S., Beckmann, C., Brammer, M., Brandeis, D., Bölte, S., Bourgeron, T., Bours, C., ... Buitelaar, J. K. (2017). The EU-AIMS Longitudinal European Autism Project (LEAP): Design and methodologies to identify and validate stratification biomarkers for autism spectrum disorders. *Molecular Autism*, 8, 24. <https://doi.org/10.1186/s13229-017-0146-8>

McGoey, K. E., DuPaul, G. J., Haley, E., & Shelton, T. L. (2007). Parent and teacher ratings of attention-deficit/hyperactivity disorder in preschool: The ADHD rating scale-IV preschool version. *Journal of Psychopathology and Behavioral Assessment*, 29(4), 269–276.

Mullen, E. M. (1995). *Mullen Scales of Early Learning*. American Guidance Services.

Nelson, L., Ye, H., Schwenn, A., Lee, S., Arabi, S., & Hutchins, B. I. (2022). Robustness of evidence reported in preprints during peer review. *The Lancet Global Health*, 10(11), e1684–e1687. [https://doi.org/10.1016/S2214-109X\(22\)00368-0](https://doi.org/10.1016/S2214-109X(22)00368-0)

Owens, J. A., Spirito, A., & Nobile, C. (2000). The Children's Sleep Habits Questionnaire (CSHQ). *Sleep*, 22(5), 517–521.

Pluess, M., Assary, E., Lionetti, F., Lester, K. J., Krapohl, E., Aron, E. N., & Aron, A. (2018). Environmental sensitivity in children: Development of the Highly Sensitive Child Scale and identification of sensitivity groups. *Developmental Psychology*, 54(1), 51–70.

Raven, J. C. (1938). Raven standard progressive matrices.

Ravens-Sieberger, U., & Bullinger, M. (1998). Assessing health-related quality of life in chronically ill children with the German KINDL: first psychometric and content analytical results. *Quality of Life Research*, 7(5), 399–407.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3-7 years: The Children's Behavior Questionnaire. *Child Development*.

Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. (2016). Vineland Adaptive Behavior Scales, Third Edition ("Vineland-3").

Sytsma, S. E., Kelley, M. L., & Wymer, J. H. (2001). Development and Initial Validation of the Child Routines Inventory. *Journal of Psychopathology and Behavioral Assessment*, 23, 241–251.

Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: A comparative study using the short sensory profile. *The American Journal of Occupational Therapy*, 61(2), 190–200.

Wilkinson, M., Dumontier, M., Aalbersberg, I., & et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>