 AnIML >

Introduction to AnIML: What It Is and Where We're At

</  AnIML >

Gary W. Kramer

Biochemical Science Division
National Institute of Standards and Technology

NIST

This presentation will:

- Discuss Why AnIML was Created
- Describe What AnIML Is
- Describe the Components of AnIML
- Describe How AnIML Works
- Discuss the Current State of AnIML

NIST

Folks, we got trouble, right here in Science City. It starts with T, and it rhymes with D, and it stands for Data…

- Can't move it
  - From instrument to instrument
  - From instrument to application
  - From application to application
- Can't interconvert it
- Can't find all its parts
- Can't look at it easily
- Can't use it with modern computing and networking technologies

NIST

There is crisis in archiving and retrieving data

- We're drowning in data, yet increasingly we cannot find our stuff because of
  - Incredible Success of Hyphenated and Multiplexed Analytical Techniques
  - Moore's Law
- Data mining increases value of archived data
- Regulatory agencies are now demanding extraordinarily long data retention
- In terms of retrieving archived data, we were better off with paper forty years ago before the arrival of lab computers…
- We can still read the data in Newton's notebooks today; will folks be able to read ours in 100, 20, or even 5 years?

NIST

The Analytical Information Markup Language (AnIML)

- has been developed expressly for analytical chemistry result data and metadata
- defines a data format for representing all types of result data and metadata
- provides a system for identifying content (data elements and attributes) with labels (tags)
- provides a model and the means for defining the structure, content, and semantics of result data in documents
- expresses shared vocabularies
- allows machines to carry out rules made by people

NIST

Key Features of AnIML

- Extensible
- Flexible - Can Accommodate ANY Type of Data
- Validateable
- Traceable through Audit Trails
- Verifiable Using Digital Signatures

NIST

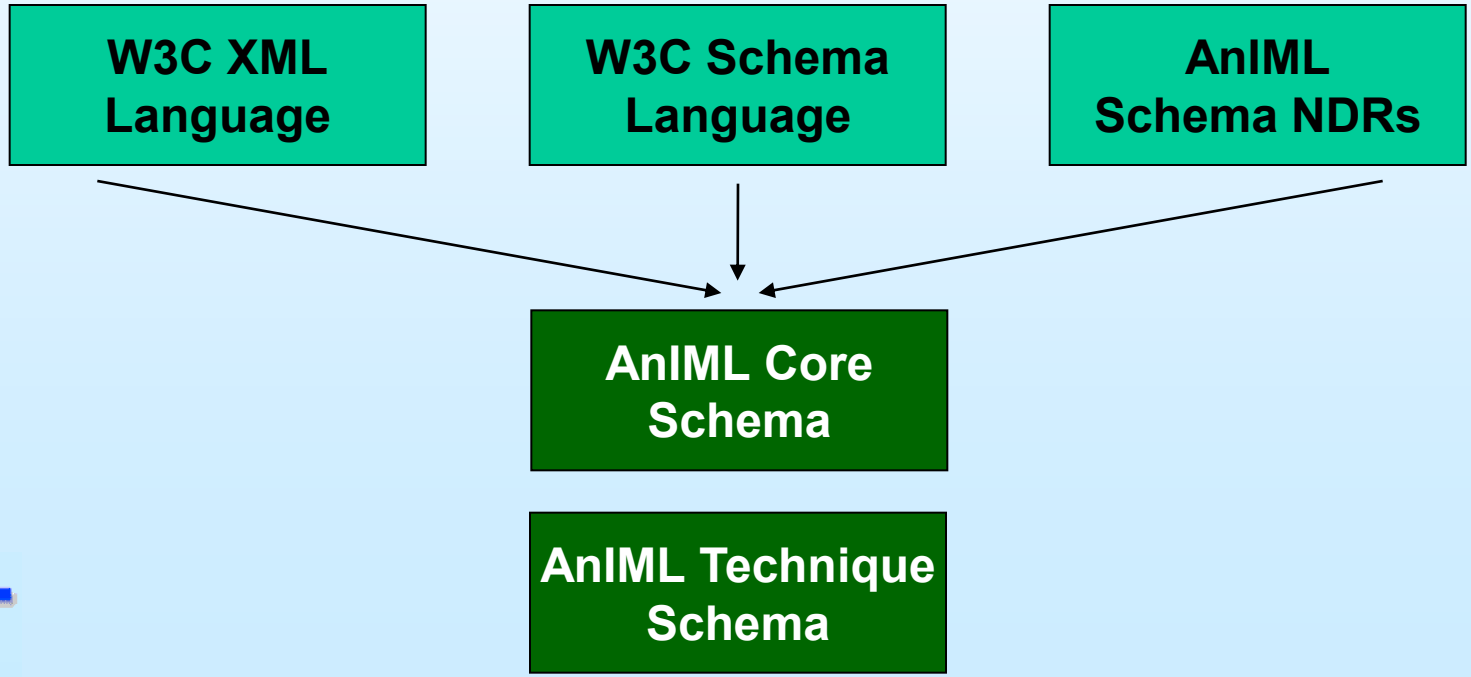
AnIML Components

- AnIML Core Schema
  - One Schema
  - Maintained by ASTM E13.15
- AnIML Technique Schema
  - One Schema
  - Maintained by ASTM E13.15
- AnIML Technique Definition Documents
  - One or More Instance Documents per Technique
  - Maintained by ASTM E13 or Appropriate Domain Expert Organization
- AnIML Technique Definition Extensions
  - One Instance Document per Technique Created from the Appropriate Base Technique Definition
  - Maintained by Vendor, Organization, User, or Whomever Extends the Technique
- AnIML Result Data Files

NIST

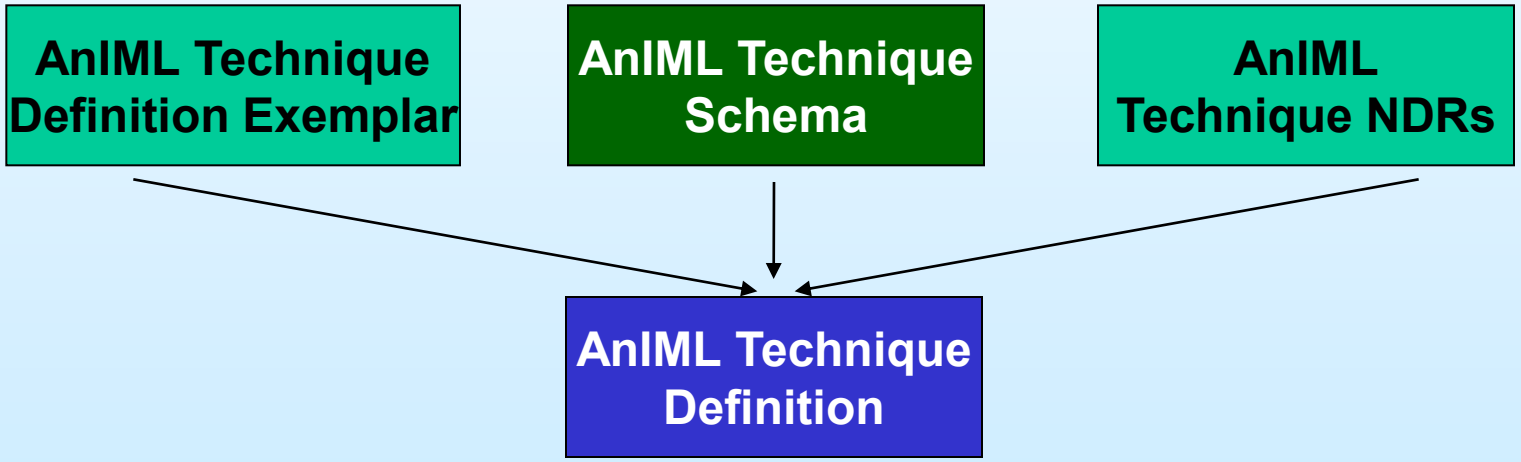
AnIML Schema

- Core Schema – a generic data content model useable for describing all/any analytical chemistry data and metadata
- Technique Schema – a content model for descriptors used to constrain the Core Schema data model to comply with the commonly accepted conventions of a specific analytical technique



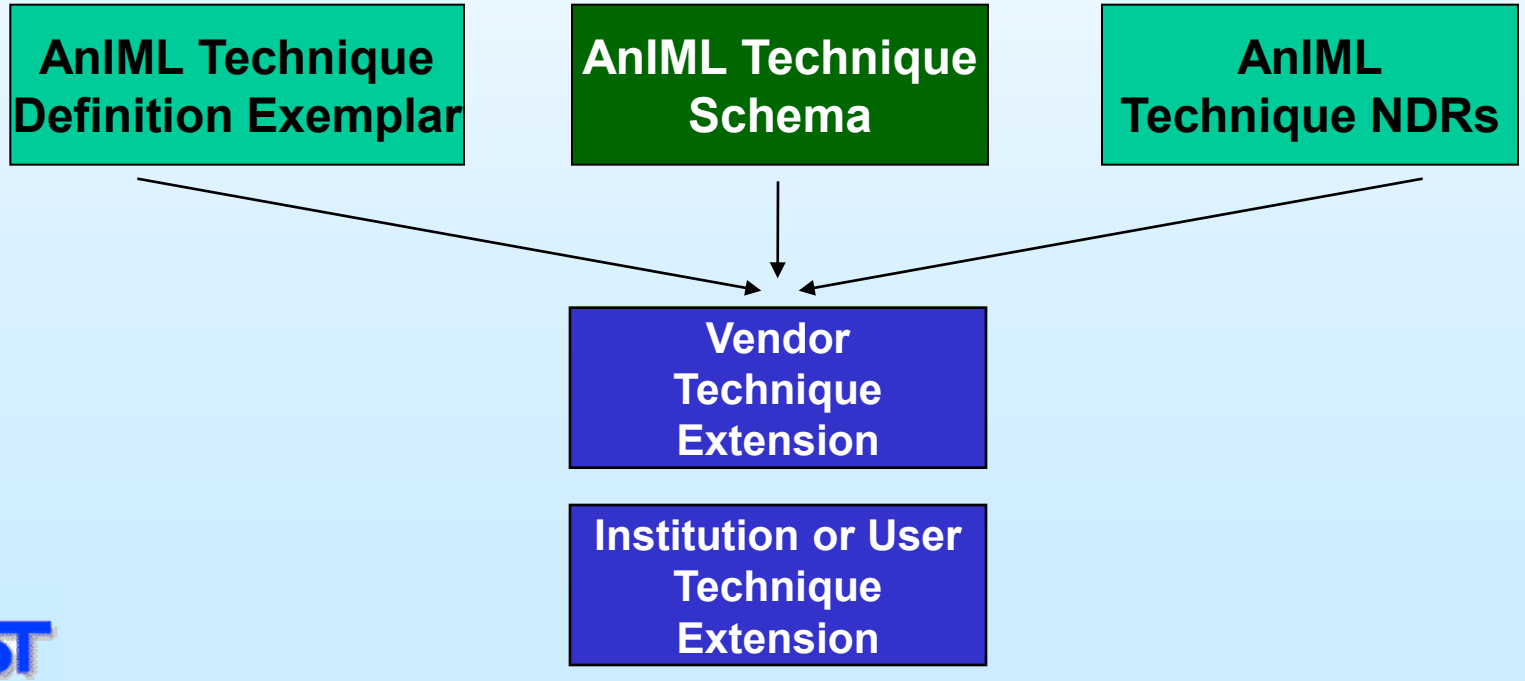
AnIML Technique Definitions

- XML Instance Documents
- Conform to the AnIML Technique Schema Data Model
- Purpose is to constrain the data representations in the very flexible AnIML Core to comply with the data formats and metadata conventions commonly accepted by those practicing the technique



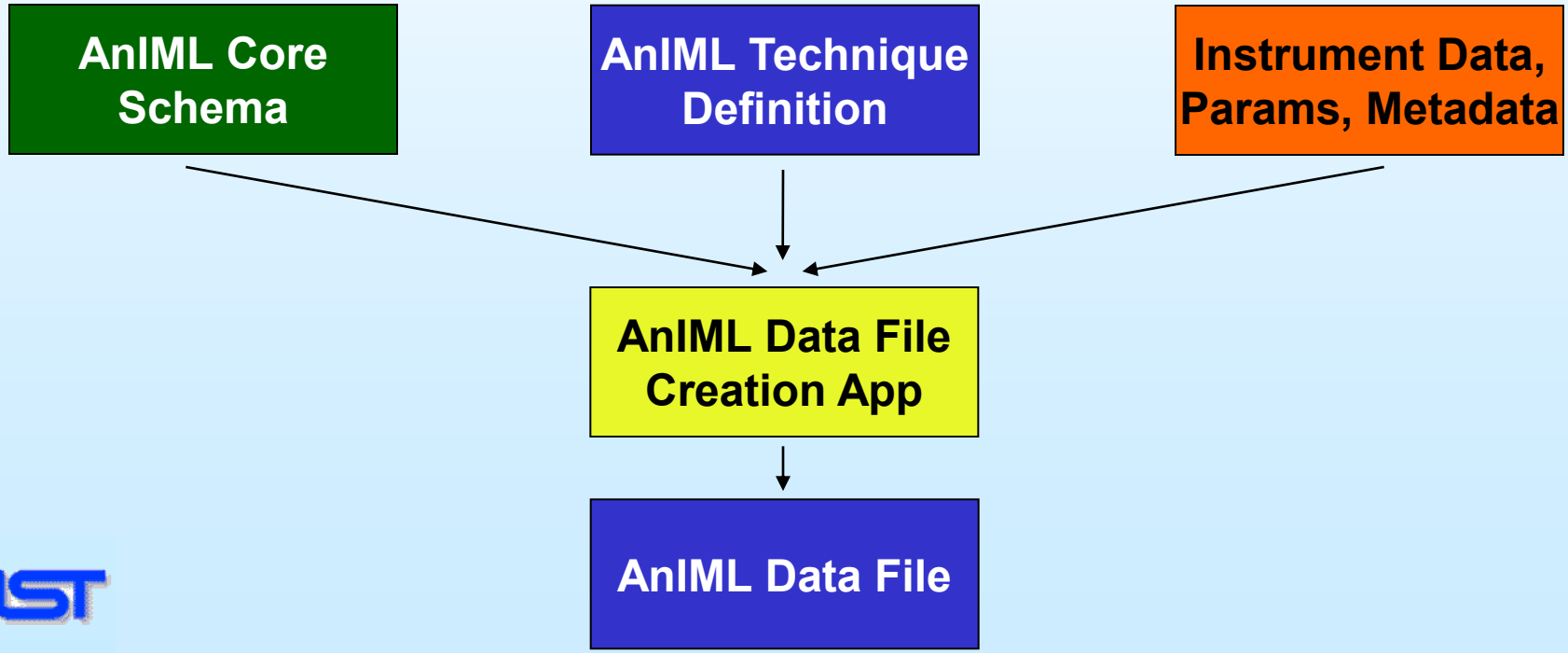
AnIML Technique Definition Extensions

- XML Instance Documents
- Conform to the AnIML Technique Schema Data Model
- Append AnIML Technique Definitions to add new data representations and metadata (but not to redefine or remove existing elements and attributes of the model)
- Provided by vendors, institutions, users…



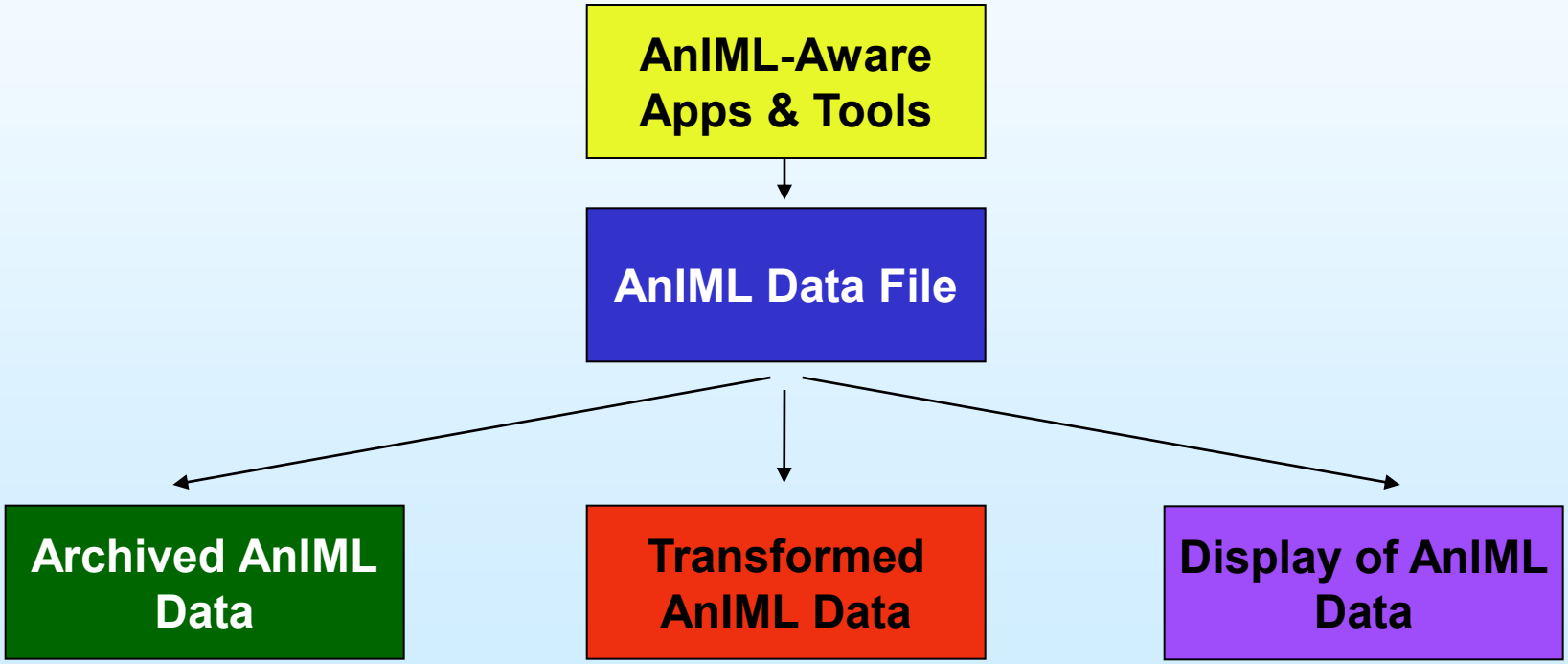
AnIML Data Files

- XML Instance Documents
- Content model defined by the AnIML Core Schema as constrained by applicable AnIML Technique Definitions
- Contains the data and metadata from the experiment, measurement, simulation, and/or data processing



AnIML-Aware Applications and Tools

- Read and parse AnIML Data Files
- View, transform, archive…AnIML Data Files
- Validate AnIML Data Files to ensure compliance



AnIML Core Data Representation

- Variable or Axis-Centric Approach
  - Variables
    - Independent Indexing
    - Independent
    - Dependent
  - Peak Tables
- Data Continuity Types
  - Continuous (Sampled)
  - Discrete
  - Sparse
- Name-Value Pairs
- IEEE 32 Floating Point Representation
- IEEE 64 Floating Point Representation
- Base 64 Encoding

NIST

AnIML Technique Definition Types

- Sample Processing
- Analyte Detecting
- Data Processing
- Process Organizing
- Data Organizing/Mapping

Prerequisites for Creating a Technique Definition

- Must have defined, agreed upon ontology
- Must have defined, agreed upon data dictionary

NIST

Sample Processing Technique Definitions

Chromatography

- Used for GC, LC, Ion, FIA…
- Does not specify detector data (i.e., no chromatograms)
- Provides for storage of flow data, temperature data, pressure data, mobile phase composition data, if present
- Defines information elements for column parameters, mobile phase parameters, injection parameters, flow rate parameters…

NIST

Analyte Detecting Technique Definitions

- UV/Vis
  - Scanning Instruments, Diode-Array Instruments, Single Wavelength, Multi Wavelength, and Diode-Array Detectors, and FT UV/Vis Instruments
- IR
  - Scanning Instruments, FT IR Instruments, Array-Detector Instruments, NIR
- MS
  - Sector-Based Instruments, TOF, Single and Multi-Quadrapole Instruments, and FT MS Instruments with a wide array of injection techniques
- 1D NMR
  - CW Instruments, FT NMR Instruments

NIST

Analyte Detecting Technique Definitions

- Point Detectors
  - FID
  - TCD
  - ECD
  - NPD
  - ELS
  - RI
  - Conductivity
  - …

Data Processing Technique Definitions

- Post Processing
  - Integration, Baseline correction, Background correction (subtraction), Reference correction (ratioing), Filtering/smoothing, Differentiation
  - Trace forming – peak identification
  - Cross-cutting data arrays for single/multiple ion or wavelength chromatograms or total ion (TIC) or total wavelength chromatograms
  - Peak table formation

NIST

Process Organizing Technique Definitions

- Indexing
  - Kinetics experiments
  - Temperature, pressure, pH… profile experiments

Data Organizing/Mapping Technique Definitions

- SEDD Stage 4

AnIML: Where We're At for Version 1.0

- AnIML Core Schema
  - Complete
  - Frozen
  - Internally Documented
- AnIML Schema NDRs
  - Complete
- AnIML Technique Schema
  - Complete
  - Frozen
  - Internally Documented
- AnIML Technique NDRs
  - In progress
- AnIML Technique Exemplar
  - In Progress

NIST

AnIML: Where We're At for Version 1.0

- AnIML Technique Definitions
  - UV-Vis Technique Definition - Complete and in Review
  - Chromatography Technique Definition - Started
  - Mass Spec Technique Definition - Started
  - 1D NMR Technique Definition – Not Started
  - IR Technique Definition – Not Started
  - Point Detector Technique Definitions – Not Started
  - AnIML Misc. Technique Definitions
    - Indexing – Complete
    - Cross-Cutting – Not Started
    - Trace Forming – Not Started
    - Peak Table – Not Started
  - SEDD – Demo Complete for GC MS Data

NIST

AnIML: Where We're At for Version 1.0

- AnIML Applications and Tools
  - AnIML Data Viewers – Several Complete
  - AnIML Extensible Validator – Complete
  - Simple AnIML File Writer for Agilent 8453 UV/Vis - Complete
  - JCAMP-DX to AnIML Data Converters – Several Complete
  - XSLT Documentation Extractor for Schemas – Complete
  - Documentation Extractor for Technique Definitions - Complete

NIST

AnIML Documents and Standards

- AnIML Standard Practice
- AnIML Core Schema and Technique Schema
  - Standard Specification
  - Standard Guide
- AnIML Technique Definition Documents
  - Standard Specification
  - Standard Guide
  - AnIML UV-Vis Technique Specification (Guide ?)
  - AnIML IR Technique Specification (Guide ?)
  - AnIML Chromatography Technique Specification (Guide ?)
  - AnIML Mass Spec Technique Specification (Guide ?)
  - AnIML 1D NMR Specification (Guide ?)
  - AnIML Misc. Technique Specifications (Guide ?)
    - Indexing
    - Point Detectors
    - Cross-Cutting
    - Trace Forming
    - Peak Table
- NIST Interagency Report

Some Current AnIML Activities

- AnIML – SEDD
  - Interaction with US EPA and US ACE
- Plan to Expedite the Completion of AnIML 1.0
  - Volunteer effort is too slow
  - There is still too much to do
  - Need paid help to develop components and documentation
  - Funded by instrument vendors and large institutional users
  - Administered by ASTM
- Project to Create AnIML File Writers for Several UV/Vis Instruments

NIST

More Information

- AnIML
  - http://animl.sourceforge.net
  - http://www.animl.org
  - http://www.iupac.org/standing/cpep.html

NIST

Acknowledgements

Mark Bean - GlaxoSmithKline
Maren Fiege - Waters Informatics GmbH
Peter Linstrom - NIST
Jamie McQuay - Scimatic Software
Dale O'Neill – Agilent Technologies
Burkhard Schäfer – BSSN

Tony Davies – IUPAC Subcommittee on Electronic Data Standards

Reinhold Schäfer – Fachhochschule Wiesbaden

- Alex Rühl
- Martin Peschke
- Aykut Arslan
- Dominick Pötz
- Anh Dao Nguyen
- Alex Roth
- Ronnie Jopp
- Patrick Gleichmann
- Kordian Placzek
- Frank Masur
- Dennis Backhaus