PROPERTY GRAPH STANDARDS, PROCESS & TIMING

Keith W. Hare

   Convenor, ISO/IEC JTC1 SC32 WG3 Database Languages

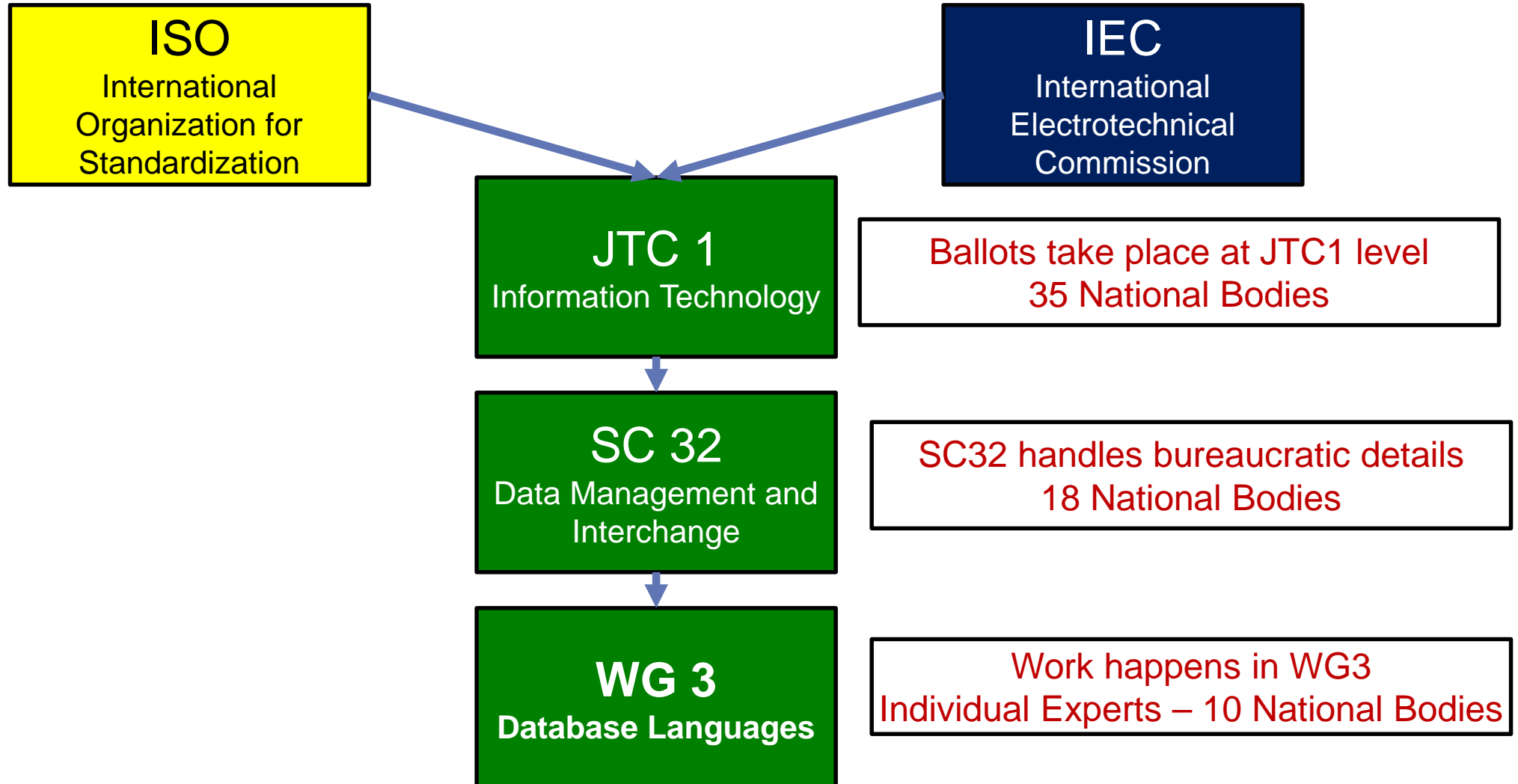June 2, 2022

Introduction

- Database Language Standards Structure and Process
  - International Standards Hierarchy
  - ISO/IEC JTC1 Standardization Process
  - USA Standards Structure
  - Who participates?
- Property Graph Standards – SQL/PGQ & GQL
  - Brief History of SQL and GQL
  - GQL Committee Draft Ballot
  - SQL/PGQ & GQL expected dates
- Summary

Database Language Standards Working Group

ISO/IEC JTC1 SC32 WG3 Database Languages

- ISO – International Organization for Standardization
- IEC – International Electrotechnical Commission
- JTC 1 – Joint Technical Committee 1 – Information Technology standards
- SC 32 – Sub Committee 32 – Data Management and Interchange
- WG 3 – Working Group 3 – Database Languages

International Standards Hierarchy

ISO
International Organization for Standardization

IEC
International Electrotechnical Commission

JTC 1
Information Technology

Ballots take place at JTC1 level
35 National Bodies

SC 32
Data Management and Interchange

SC32 handles bureaucratic details
18 National Bodies

WG 3
Database Languages

Work happens in WG3
Individual Experts – 10 National Bodies

Standardization Steps and Acronyms

- New Work Item Proposal – NWIP

- Working Draft – WD

- Committee Draft – CD

- Draft International Standard – DIS

- Final Draft International Standard – FDIS

- International Standard – IS

ISO/IEC JTC1 Standardization Process
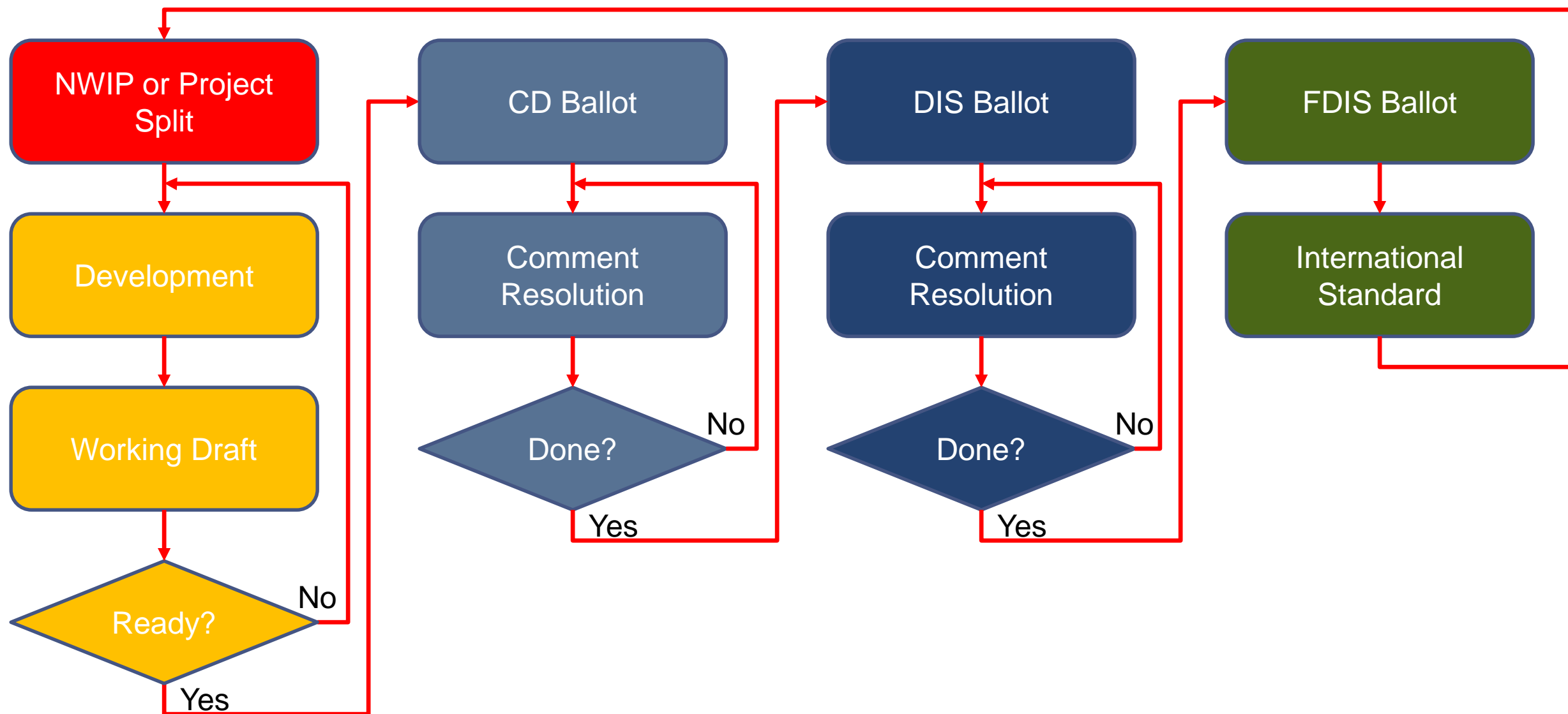
Who participates – SC32 WG3?

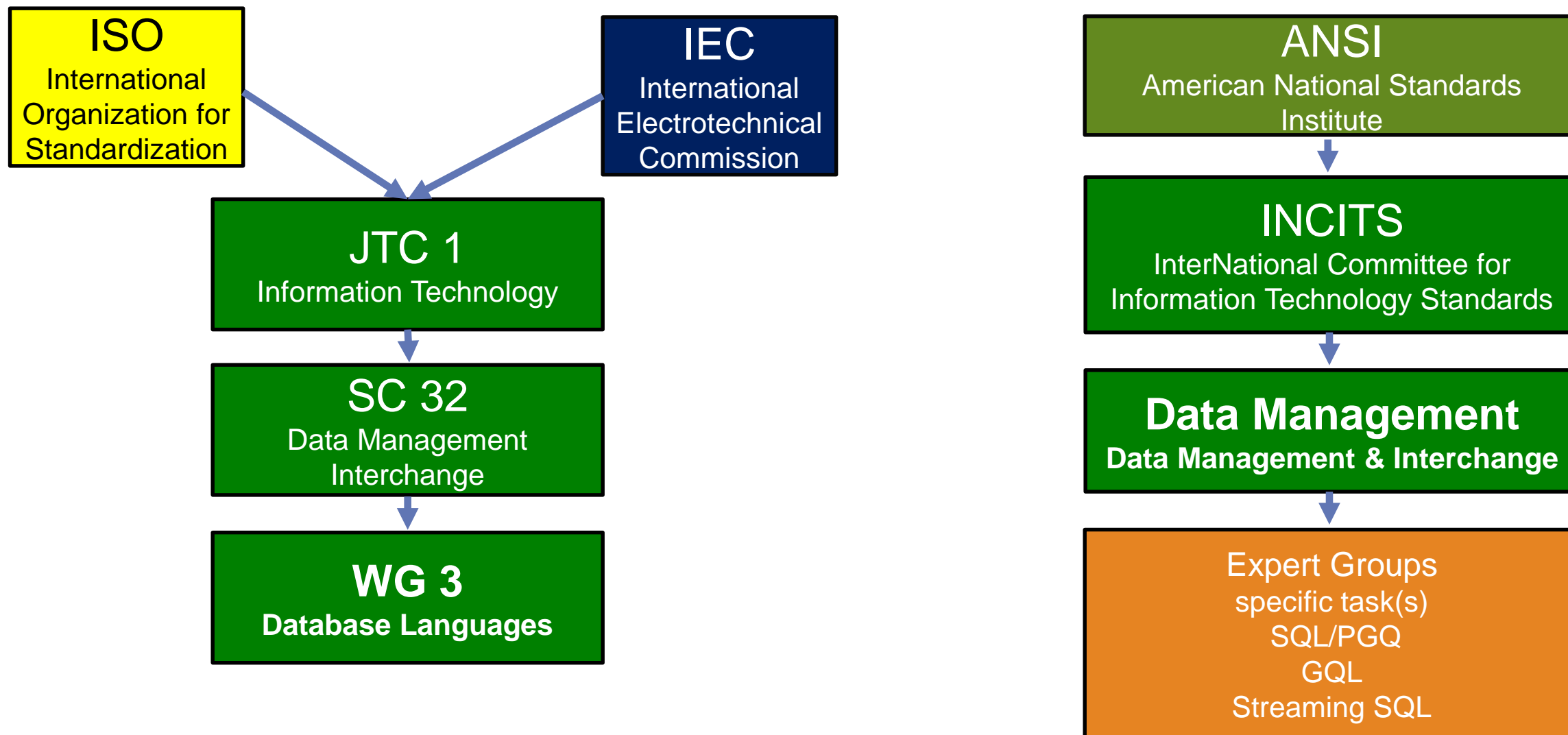Experts from the following national bodies participate in SC32 WG3:

1. China
2. Denmark
3. Finland
4. Germany
5. Japan
6. Korea
7. Netherlands
8. Sweden
9. United Kingdom
10. United States

Organizations Participating in National Bodies

- China
  - Ant Financial
  - Boray Data
  - CESI
  - Huawei
- Denmark
  - TF Informatik
- Finland
  - Profium
- Germany
  - EDB
  - Oracle

- Japan
  - Hitachi
  - Tokyo Metropolitan University
- Korea
  - Bundang Hospital
  - CnTechSystems
- Netherlands
  - Cannan Consultancy
- Sweden
  - Neo4j
- United Kingdom
  - PR Brown
  - University of Edinburgh
- USA (see later slide)

Note: This list is probably incomplete

International Hierarchy mirrored in the US

ISO
International Organization for Standardization

IEC
International Electrotechnical Commission

JTC 1
Information Technology

SC 32
Data Management Interchange

WG 3
Database Languages

ANSI
American National Standards Institute

INCITS
InterNational Committee for Information Technology Standards

Data Management
Data Management & Interchange

Expert Groups
specific task(s)
SQL/PGQ
GQL
Streaming SQL

INCITS Data Management (USA) Participants

Mostly SQL
- Actian Corporation
- IBM Corporation
- Intersystems Corporation
- Microsoft Corporation
- Oracle
- SAP
- Teradata

Mostly GQL
- ArangoDB Inc
- FairCom USA
- Google
- Intel
- JCC Consulting Inc
- Katana Graph
- Neo4j Inc
- Optum Technology
- Redis Labs
- TigerGraph

Mostly Streaming SQL
- Alibaba Group
- Amazon Web Services
- Boray Data
- Confluent
- Hazelcast
- Materialize
- Snowflake
- SQLstream, a Thales Company

Mostly Metadata
- Farance Inc
- Institute for Defense Analyses
- William McCarthy
- National Cancer Institute
- Nurocor

Working In INCITS Data Management

- Work done by interested parties in the Expert Groups (EG)
  - Property Graph Queries in SQL – SQL/PGQ
  - GQL
  - PGQ & GQL have 2-hour web conferences on alternate weeks
- Discussions in the EG is based on written papers
  - Concrete change proposals
  - Discussion papers
    - Basis for discussion of designs or alternatives
    - Discussion will eventually lead to a Change Proposal
  - EG agrees to submit papers to Data Management and WG3
- Data Management (formerly DM32) reviews EG contributions

Working In ISO/IEC JTC1 SC32 WG

- International Committee
  - Participants from national standards bodies
  - Participants operate as individual experts
- Web conferences or face-to-face meetings
  - Web conference times offset by 8 hours every month
  - June 13-24 2022 in Berlin – hybrid
- Written papers
  - Concrete change proposals
  - Discussion papers
- Final decisions are made in WG3
  - Decisions recorded in minutes
  - Editors apply approved papers to drafts

SC32 WG3 Formal Liaison Relationships

- LDBC (Linked Data Benchmark Council) — liaison since 2017
  - Industry/Academic consortium focused on graphs
  - Started with benchmarks & has evolved to model, language, and other topics
  - Working Groups of interest to SC32 WG3 focus on property graph language – PGQ & SQL
    - Existing Languages, Property Graph Schema, GQL Formal Semantics working groups
  - Support/strengthen WG3 standards
    - Review of WG3 documents
    - Contribution of papers to WG3 (critique/corrections, feature suggestions)
  - Property graph theory
    - ACM SIGMOD/PODS 2021 "PG-Keys: Keys for Property Graphs"
    - ACM SIGMOD/PODS 2022 "Graph Pattern Matching in GQL and SQL/PGQ"
  - An evolving bi-directional process for collaboration
- OGC (Open Geospatial Consortium)
  - Requirements for supporting spatial data in GQL (v2 or later)

SQL Standards – a brief history

ISO/IEC 9075 Database Language SQL

- SQL-87 – Transactions, Create, Read, Update, Delete
- SQL-89 – Referential Integrity
- SQL-92 – Internationalization, etc.
- SQL:1999 – User Defined Types
- SQL:2003 – XML & OLAP
- SQL:2008 – Expansions and corrections
- SQL:2011 – Temporal
- SQL:2016 – JSON, RPR, PTF, MDA (2019)
- **SQL:202x – SQL/PGQ, Property Graph Queries in SQL**
  - ISO/IEC 9075-16 *Information technology — Database languages SQL — Part 16: SQL Property Graph Queries (SQL/PGQ)*
  - SQL/PGQ project initiated September, 2017

GQL Standards – a brief history

ISO/IEC 39075 *Information Technology — Database Languages — GQL*

- Create, Read, Update, Delete, Transactions, Schema
- GQL project initiated September, 2019

Property Graphs – SQL/PGQ and GQL

SQL/PGQ

- Property Graph views of SQL tables
- **Graph Pattern Matching queries**
  - GRAPH_TABLE() in SQL FROM
  - Supports Reads
- Common foundation with SQL and graph query languages
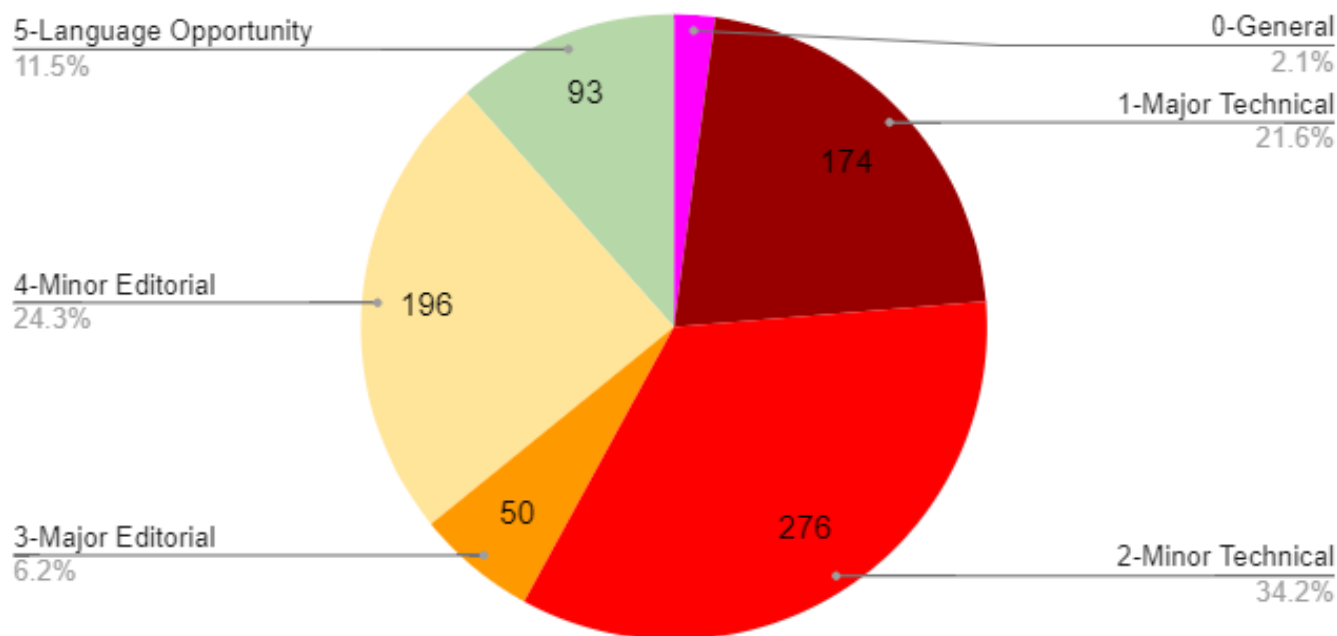- Does not support schema-flexible graphs

GQL

- Full DB language
  - DML – Create, Read, Update, Delete
  - DDL – Create Type, Create Graph
- **Graph Pattern Matching queries**
- Leverages common foundation from SQL and property graph languages
- Supports schema-fixed and schema-flexible variants

GQL Committee Draft (CD) Results

- Ballot closed February 2022, Comments more important than the ballot results

Severity of all comments

| Category	%	Count
5-Language Opportunity	11.5%	93
0-General	2.1%	
1-Major Technical	21.6%	174
4-Minor Editorial	24.3%	196
2-Minor Technical	34.2%	276
3-Major Editorial	6.2%	50

- Editors resolve most Major (50) and Minor (196) Editorial comments – 31%
- Many (most?) Language Opportunities (93) deferred to a future GQL – 11%
- 159 are Possible Problems known before the GQL CD ballot

GQL Comment Resolution Progress (2022-06-10)

Resolved/Assigned
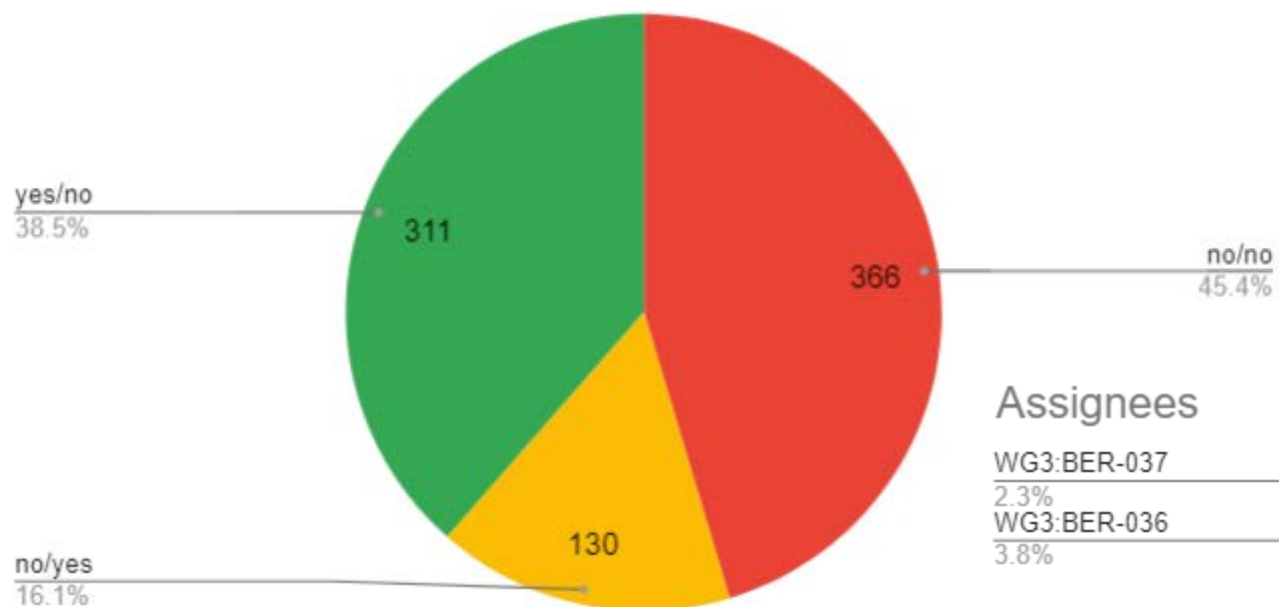
Assignees
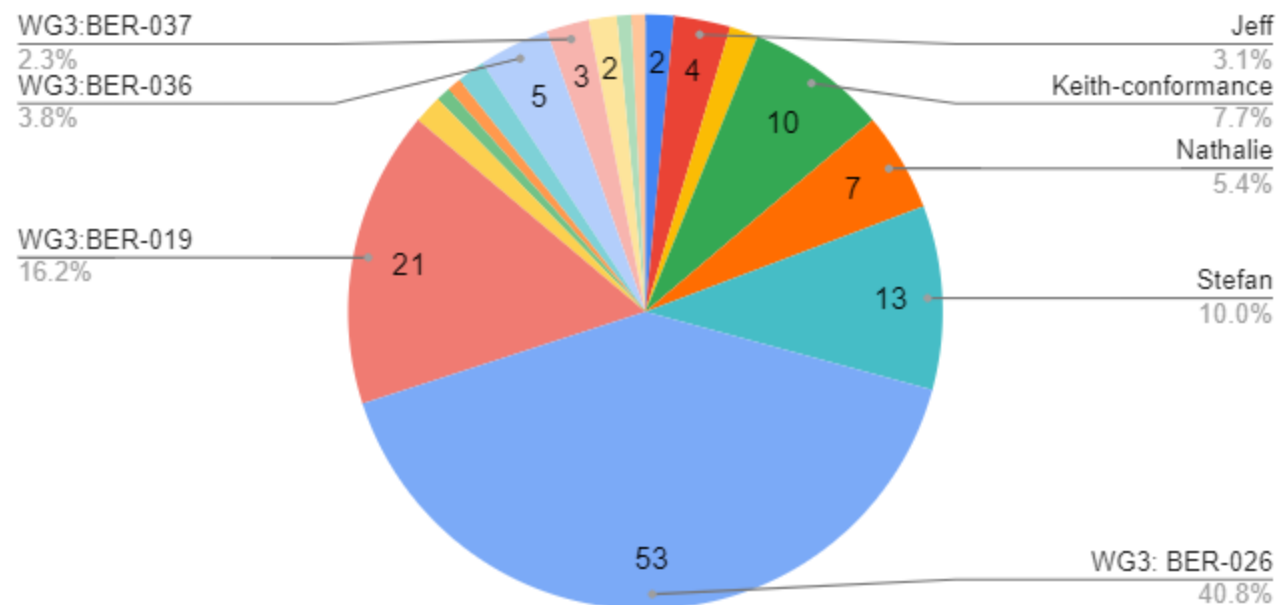
Expected Dates

- SQL/PGQ
  - CD Ballot completed – February 2021
  - **DIS Ballot starts – August 2022**
  - DIS Ballot completes – December 2022
  - Published standard – June 2023
- GQL V1
  - CD Ballot completes – February 2022
  - **DIS Ballot Starts – April 2023**
  - DIS Ballot completes – September 2023
  - Published standard – March 2024
- Draft Standards are stable by DIS ballot start

Summary

- Standards Process
  - Iterative, collaborative process
  - Some amount of standards bureaucracy
  - Tedious at times
  - Results are pretty good
- Property Graph Standards
  - SQL/PGQ published June 2023
  - GQL published March 2024
  - Graph Pattern Matching (GQL and SQL/PGQ) stable August 2022