

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
23 March 2023 (23.03.2023)



(10) International Publication Number
WO 2023/044445 A1

(51) International Patent Classification:
G06F 16/25 (2019.01)

(21) International Application Number:
PCT/US2022/076595

(22) International Filing Date:
16 September 2022 (16.09.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/245,244 17 September 2021 (17.09.2021) US
17/665,109 04 February 2022 (04.02.2022) US

(71) Applicant: **AB INITIO TECHNOLOGY LLC** [US/US];
201 Spring Street, Lexington, Massachusetts 02421 (US).

(72) Inventors: **RADIOJEVIC, Dusan**; 201 Spring Street,
Lexington, Massachusetts 02421 (US). **PARKS, Robert**;
201 Spring Street, Lexington, Massachusetts 02421 (US).
WEISS, Adam; 201 Spring Street, Lexington, Massachu-
setts 02421 (US). **JANKOVIC, Maja**; 201 Spring Street,

Lexington, Massachusetts 02421 (US). **VICKERY, John**;
201 Spring Street, Lexington, Massachusetts 02421 (US).

(74) Agent: **REILLY, Christopher L.** et al.; Fish & Richardson
P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022
(US).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ,
CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE,
KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU,
LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA,
NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,
RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH,
TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS,
ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ,
UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

(54) Title: METADATA-DRIVEN DATA INGESTION

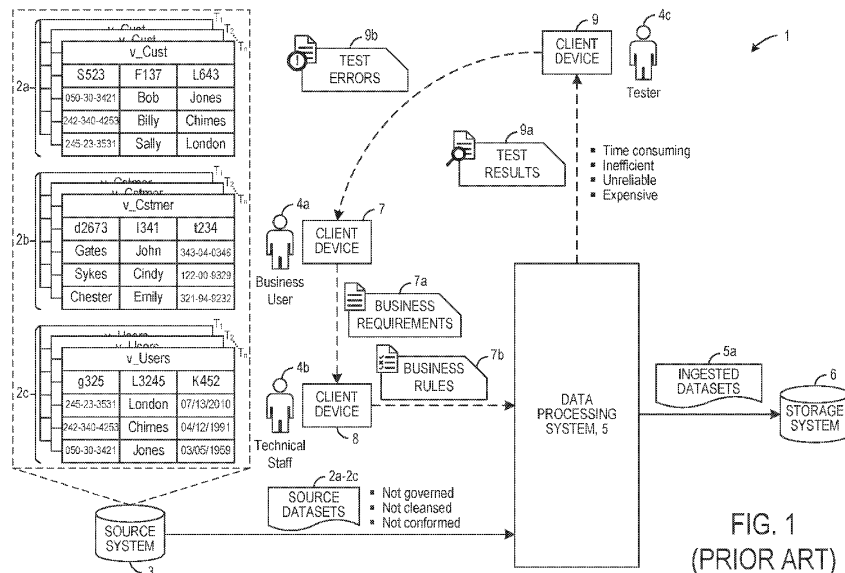


FIG. 1
(PRIOR ART)

(57) Abstract: An electronic system for increasing the speed of preparing data with a specified data quality for storage by automatically identifying for a user, with minimal user input, common contexts among (i) fields in disparate datasets, and (ii) names the user has specified as potentially describing the fields, and by using those common contexts to govern the disparate datasets prior to storage to ensure the specified data quality.

[Continued on next page]

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*