**Appendix**

**Figure A. The institutional landscape of tax fraud detection in Belgium**



**Figure B. The institutional landscape of social security fraud detection in Belgium**



Table A. Interview List

|  |  |
| --- | --- |
| **Position** | **Name and type of organization** |
| Data Miner | FPS Finance (Federal) |
| Service Manager (Data Warehouse) | FPS Finance (Federal) |
| Head of Data Miners | FPS Finance (Federal) |
| e-Auditor | FPS Finance (Federal) |
| Data Miner | FPS Finance (Federal) |
| Fiscal Coordinator | FPS Finance (Federal) |
| Regional Director | FPS Finance (Federal) |
| Operation Manager | FPS Finance (Federal) |
| Advisor - Administration générale Expertise et Support stratégiques | Centre des connaissances | FPS Finance (Federal) |
| Research Consultant | Smals (non-profit private org.) |
| Research Consultant | Smals (non-profit private org.) |
| Research Consultant | Smals (non-profit private org.) |
| Administrator | Crossroad Bank for Social Security |
| Vice-administrator | Crossroad Bank for Social Security |
| Deputy general administrator | INAMI (federal) |
| Head of Data Management | INAMI (federal) |
| Chief Data Officer | INAMI (federal) |
| Deputy general administrator | INAMI (federal) |
| Social Inspector | INAMI (federal) |
| Director of the Social Inspection Direction | INAMI (federal) |
| General Administrator | ONEM (federal) |
| IT researcher in data mining  | ONSS (federal) |
| Data scientist  | ONSS (federal) |
| Director service taxation and valuation | VLABEL (regional) |
| Employee competence centre reporting | VLABEL (regional) |
| Data protection officer | A Flemish agency in the social affairs domain, participated on the condition of anonymizing both name and organization (regional) |
| General advisor  | Agoria (private-profit) |
| Professor | UCL/ICTEAM (research institute) |
| Director | HEC Digital Lab (research institute) |
| Assistant Professor | UMons (research institute) |
| Advisor | FGTB (union) |
| Program Manager | NRB (PPP) |
| Partner | PwC (private-profit) |
| Partner | PwC (private-profit) |
| General Manager | Febetra (business federation) |
| Director Social Department | Confederatie Bouw (business federation) |
| Advisor | Confederatie Bouw (business federation) |
| General Director | Constructiv (service provider org.) |
| Former Secretary-General | MC (mutuality) |
| Head of Study Department  | MC (mutuality) |
| Advisor | MC (mutuality) |
| Information & Technology Manager | AdN (regional) |
| Expert | AdN (regional) |
| Head of Study Department | UCM (employer org.) |
| Secretary-General | UCM (employer org.) |
| Social Affairs Advisor | UCM (employer org.) |
| Social Security Advisor  | UCM (employer org.) |
| Associate Professor | HEC (research inst.) |
| Customer Solution Manager | SAS (private-profit) |
| Head of Task Department | FEB (employer org.) |
| Social Security Advisor | FEB (employer org.) |
| Data Analyst | Ecolo (political party) |
| Digital Advisor | Centre Jean Gol (political party) |
| Tax Advisor  | Centre Jean Gol (political party) |
| Secretary-General | CNCD 11 11 11 (NGO) |
| Head of Study Department | CSC (union) |
| Advisor | IEV (political party) |
| Head of Study Department | PTB (political party) |
| National Technology Officer | Microsoft (private-profit) |
| Secretary-General | UPTR (federation of companies) |
| CEO | Skwarel (start-up) |
| Director Task & Public Affairs | Febelfin (federation of companies) |
| Professor  | VUB AI Lab (research inst.) |
| President  | BATL (Professional association) |
| Advisor  | BATL (Professional association) |
| Analysis department director | CTIF (federal)  |

Table B. Codification list

|  |  |  |  |
| --- | --- | --- | --- |
| **Theoretical framework** | **Constructs** | **Elements** | **Topics** |
| Institution-driven  | Regulations | Data  | GDPR, national laws, data combination, data anonymization (difficulty of doing, and viable use of it) |
| Taxation/Social Security | national laws, EU laws, sanctions |
| Transcending laws | environmental laws |
| Justification of decisions | the challenge of explainability with deep learning |
| User-driven | Trust | Trust in administration | tax authorities, inability to develop in-house solutions, regulative agencies, trust in centralized entity vs the necessity of blockchain |
| Trust in society | no adherence to tax, easy acceptance of new technologies (not critical thinking), lack of skills regarding new technologies |
| Trust in technology | Cloud systems, social acceptability, algorithms, blockchain, political distrust, digitalization improves trusts, social distrust in new digital technologies (related to Fake news, conspiracy theories, anti-science speeches), |
| Trust in system | Surveilance capitalism, undeclared purposes (e.g. commercial, political) |
| Trust in tech providers/private sector | implementation of ethical standards, GAFAM (big tech), higher efficiency and expertise than public sector, not adapted to local challenges (better in-house), distrust to private sector in civil servants, distrust to banks to share data |
| Institution-driven | Management/operational systems | Guidelines  | Legal guidelines, a precise framework to declare fraud suspicion (for banks, casinos etc.) |
| Rules & standards | data confidentiality, ethical rules and standards, lack of standards to assess overall quality of taxation system (e.g. ISO) |
| Principles | only-once, trade-offs (weak/strong AI vs ethics, digitalization vs reduction in human resources, efficiency vs transparency, cost reduction vs quality of services, decentralization of data vs difficulty of analysis, privacy vs the cost of more advanced cytopgraphic solutions, cost of investment vs return of investment, openness vs not sharing too much information about business), lack of flexibility in project requirements  |
| Processes | risk assessment, human in/on the loop, safeguards against AI bias, AI-led data management, streamlining, process automation, reducing administrative burden, administrative simplification (e.g. public procurement), automation of rights, elimination of unnecessary tasks, reporting obligations for companies, treating big companies same as small companies, follow-up after the detection of tax fraud, data sorting, data protection, very long specific investigations vs agile way of working (thanks to data insights), data matching and cross-checking across institutions, purpose delienation, making proportionate cross-checks, administrative burden on authorization of projects (delays timely detection of new types of fraud), public registry for used algorithms (e.g. Amsterdam algorithm register) |
| Strategies | data strategy, 5-10 year digital investment plans, a central data management political vision, a common data collection and exploitation strategy, horizontal control, use of AI and data on most problematic cases, big fish vs small fish, incentives to entrepreneurs, pre-filling of tax returns, specifities of SMEs, improving targeting of controls (focus on fraudsters leave others alone), early detection of frauds, creation of technology watchgroups in administrations (to improve anticipation), subscription-based IT systems, new business models for media to inform society, operational strategy on data literacy, starting small with a few tenants on big data and AI solutions and incremental increase in time with trusted tenants |
| User-driven | Perceived risk | Legal challenge | protection of citizen rights, implications to tax lawyers, rules on data confidentiality, evolution of identity |
| Control of data | tech vs government |
| Democratic challenge | autoritarian drift, development of a public GAFAM to preserve democracy and equality, avoiding Chinese system |
| Administrative challenges | disrupting administrative practices and culture, the need of reinventing itself, difficult to keep public services relevant, budget cuts and personal reduction will human inspectors will not have time to verify AI's decisions |
| Societal challenges | job loss, decrease in working time, disruption in labour market (rapidly changing tech skills), environmental impact, increasing digital divide |
| Institution-driven | Governance system | Data governance | sensitive data, reliability of databases (e.g. updating, info about foreign workers), secure networks to exchange data with mutualiteit, collaboration between data centres, exchanges between databases (e.g. social security data+financial data), siloed organizations, automatic collection of transaction data at transaction level, data sharing with banks, access and control of data, inclusivity, frequency of checks on data quality, internal audit processes (e.g. data anonymization) |
| Open governance  | open data, open sources, better control for experienced inviduals and groups, giving citizens access, reduced control and access to software, access to data by the private sector, use of open data to better understand the tool, user friendliness of the interface  |
| Multi-level governance | global data platform for tax authorities, international coordination at the EU level, social dumping, collaboration between Belgian and foreign supervisory bodies, administrative fragmentation (local-regional-federal), institutional complexity, EU as a normative actor (e.g. green deal, free movement of workers), EU as a coordinator |
| Network governance | institutional rivalry, coordination between services, collaboration with companies and citizens, expertise centres, collaboration with other public actors (police, social security, justice), understanding the challenges of entrepreneurs, need for listening many actors (including private actors and NGOs), control of compliance, finding innovation champions |
| Institution-driven  | Technical infrastructure | Security | data security, network security, data privacy |
| Quality of database | social security data about foreign workers, up-to-dateness of data |
| Data collection & analysis | Use of AI with other tools (e.g. IOT), blockchain/DLT, data hubs (across sector, across countries), automation, predictive analysis, setting algorithms properly (parameter choices, big vs small fish), data mining, sensitivity of data, classification of data, probabilistic vs deterministic models, open data, big data platform |
| Softwares | subscription-based IT systems, supply model of software, open vs proprietary softwares, entanglement with obsolete legacy software |
| Computer maturity |  |
| Reliance/dependence on external actors | GAFAM, SAS, in-house solution vs outsourcing |
| User- and institution-driven | Public values | Appropriateness of technology | why we use AI, avoiding discrimination, asking ethical/moral questions, avoid nervousness of an association with “Big Brother”.  |
| Respecting privacy | profiling fraudsters |
| Tax fairness |  |
| Institution-driven (objective information technologies) | Technological maturity | Bias and noise | risk of bias with AI algorithm, misinterpretation, possible errors in probabilistic AI, systemic biases due to algorithm, bias in training data, using too many algorithms |
| Technology convergence | blockchain and AI, AI and IoT, machine learning for risk assessment of machine learning, BCT as a means of transparency and trust |
| Blockchain/DLT | not reached to maturity, disruptive potential, difficult to find good use cases, GDPR is an obstacle |
| AI/ machine learning | absence of AI use cases that are effective as current data-matching operations |
| Fraud detection technologies | fraud analytics, predictive models, proactive/anticipatory use of ML, tools for better targeting and sorting data, nowcasting tools |
| Institution-driven | Interoperability | Technical interoperability | lack of harmonization of IT systems in administration, compatibility of systems and means of data exchange between administrations |
| Semantic interoperability | homogenization of data, lack of standardization to ensure data quality, cross-border differences in the use of data and metadata (e.g. easier to do with some countries (Netherlands) than others) |
| Organizational interoperability | coordination between services, a harmonized social security database at the European level, aligning data strategies of multiple actors involved in social security, different applications used by insurance companies (a need for a common application for coordination with the RIZIV) |
| Regulative interoperability | data exchange between EU countries about workers, EU as a coordinator (e.g. European Labour Authority) |
| Institution- and user-driven | Capacities, skills and competencies | Resources | human resources (not enough manpower to process data, growing need to IT skills), financial resources, lack of resources at SMEs for digital solutions/sharing faster and better data with tax authorities |
| Digital skills | knowledge of algorithms and BCT, data management, critical thinking, data literacy, data storage and sharing, lack of digital skills in society, lack of digital skills in unions, lack of digital skills for new technologies in companies/SMEs |
| Training | problem of initial training of civil servants, lack of competence and poor training in tax administration, technological developments, foresight on technological changes, access to appropriate IT training, rapid change of technological skills, development of civic and professional skills in public, inclusivity of training (including migrants), training citizens about data protection, information campaigns for citizens and public agents |
| Institution-driven | Policy priorities | EU-level policy priorities | Competion with US, and Asian tech providers (EU wants to lead in producing data standards because it missed out digital transition), Green deal, Digital innovation |
| Fight against fraud | Higher priority of figthing tax fraud against social fraud, targeting small business vs big business, too much time and money in fighting social fraud less effectiveness, ineffectiveness of tax policies against large players, finding a balance between fight against fraud and protection of individual liberties, more targeted and proportional measures against tax fraud, use of data and AI to select more problematic cases, government wants total transparency on assets for tax justice, but going other way, the need for a clear vision and consistency in targeting big fraudsters in specific sectors, impression of hunt for money than fraud  |
| Political support | Need for political support for data sharing, budget allocation, short-termism |
| Geo-political aspects | Rare metals and potential conflicts, impact of technology on environment |
| User-driven | Perceived usefulness | Automation | to free up some working time for other tasks, improves efficiency, elimination of unnecessary tasks, reduces administrative costs and pushes for simplification, it saves money and brain power that can be used to better look at data and make better decision |
| Social security and taxation | for the public sector, for enterprises, for the funding of social security, citizens do not need to go administration, gain speed and efficiency to target and control problematic companies; better target controls, data matching helps to thwart the creativity of fraudsters and allows to detect low-level social fraud that is generally difficult to identify, not only detecting undue payments but ensuring people receive everything they need on time |
| Data collection & analysis | the opportunity to automate and reduce the administrative burden while improving services by anticipating problems (e.g. in mutualiteits), solving problems not being able to solve before; anticipating future problems, understanding better the reasons behind succesful practices; better results with fewer and better qualified people, making more effective policies; new IT tools accelerate data analysis while reducing the risk of committing mistakes |
| Past experiences | CBSS improved results on fight against social fraud, pre-filling of tax returns , usefulness of relational business data (SQL) rather than big data for fraud detection (ONSS), PoCs of predictive analysis with anonymised data were useful for inspectors, and that there was an enormous gain in efficiency and accuracy in terms of case selection and they wielded larger results. |
| Indirect added value of new digital technologies | Better predictive medicine impacts social security, fighting against unfair competition, making public authorities legimate and relevant |
| Institution-and user-driven | Socio-cultural elements | Digital culture | Lack of digital and data culture in administration, lack of digital culture in society, different digital mindset between Wallonia and Flanders (data protection vs innovation with SMEs) |
| Digital divide | Generational gap, socio-economic gap, gap between SMEs, digitalization of society, access to Internet,  |
| Willigness to share data |  |

Table C. Initial Reachability Matrix (IRM)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1R [ i, j] | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 |
| 2 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 3 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 4 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |
| 5 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 6 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| 7 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 |
| 9 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 10 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 12 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 |
| 13 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

Table D. Final Reachability Matrix (FRM)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| VR | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | DRP |  |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 12 |  |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 12 |  |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 12 |  |
| 7 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 8 |  |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 12 |  |
| 10 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 12 |  |
| 11 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 7 |  |
| 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| 13 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 13 |  |
| DEP | 11 | 12 | 11 | 13 | 13 | 13 | 13 | 11 | 13 | 12 | 12 | 11 | 8 | **153** |  |
| 1. Technological maturity, 2. Perceived usefulness, 3. Capacities, skills, and competencies, 4. Management/operational systems, 5. Perceived risk, 6. Governance system, 7. Technical infrastructure, 8. Public values, 9. Trust, 10. Sociocultural elements, 11. Interoperability, 12. Policy priorities, 13. Regulations, VR: variable; DEP: Dependence Power; DRP: Driving Power |
|  |
|  |

Table E. Level Partitions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Iteration 1 |  |  |  |
| **Element *P*(*i*)** | **Reachability set: *R*(*Pi*)** | **Antecedent set: *A*(*Pi*)** | **Intersection set: *R*(*Pi*) ∩ *A*(*Pi*)** | **Level** |
| 1 | 1,2,3,4,5,6,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,8,9,10,12,13 | 1 |
| 2 | 1,2,3,4,5,6,7,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,11,12,13 | 1,2,3,4,5,6,7,8,9,12,13 |   |
| 3 | 1,2,3,4,5,6,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,8,9,10,12,13 | 1 |
| 4 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12 | 1,2,3,4,5,6,7,8,9,10,11,12 |   |
| 5 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1 |
| 6 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12 | 1,2,3,4,5,6,7,8,9,10,11,12 |   |
| 7 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 2,4,5,6,7,9,10,11 | 2,4,5,6,7,9,10,11 |   |
| 8 | 1,2,3,4,5,6,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,8,9,10,12,13 | 1 |
| 9 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,10,12,13 |   |
| 10 | 1,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12 | 1,3,4,5,6,7,8,9,10,11,12 |   |
| 11 | 1,2,3,4,5,6,7,8,10,11,12,13 | 4,5,6,7,9,10,11 | 4,5,6,7,10,11 |   |
| 12 | 1,2,3,4,5,6,8,9,10,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,4,5,6,8,9,10,12,13 | 1 |
| 13 | 1,2,3,5,8,9,12,13 | 1,2,3,4,5,6,7,8,9,10,11,12,13 | 1,2,3,5,8,9,12,13 | 1 |
|  |  |  |  |  |
|  |  |  |  |  |
|  | Iteration 2 |  |  |  |
| **Element *P*(*i*)** | **Reachability set: *R*(*Pi*)** | **Antecedent set: *A*(*Pi*)** | **Intersection set: *R*(*Pi*) ∩ *A*(*Pi*)** | **Level** |
| 2 | 2,4,6,7,9,10 | 2,4,6,7,9,11 | 2,4,6,7,9 |   |
| 4 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2 |
| 6 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2 |
| 7 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 2 |
| 9 | 2,4,6,7,9,10,11 | 2,4,6,7,9,10 | 2,4,6,7,9,10 |   |
| 10 | 4,6,7,9,10,11 | 2,4,6,7,9,10,11 | 4,6,7,9,10,11 | 2 |
| 11 | 2,4,6,7,10,11 | 4,6,7,9,10,11 | 4,6,7,9,10,11 |   |
|  |  |  |  |  |
|  | Iteration 3 |  |  |  |
| **Element *P*(*i*)** | **Reachability set: *R*(*Pi*)** | **Antecedent set: *A*(*Pi*)** | **Intersection set: *R*(*Pi*) ∩ *A*(*Pi*)** | **Level** |
| 2 | 2,9 | 2,9,11 | 2,9 | 3 |
| 9 | 2,9,11 | 2,9 | 2,9 |   |
| 11 | 2,11 | 9,11 | 9,11 |   |
|  |  |  |  |  |
|  | Iteration 4 |  |  |  |
| **Element *P*(*i*)** | **Reachability set: *R*(*Pi*)** | **Antecedent set: *A*(*Pi*)** | **Intersection set: *R*(*Pi*) ∩ *A*(*Pi*)** | **Level** |
| 9 | 9,11 | 9 | 9 |   |
| 11 | 11 | 9,11 | 11 | 4 |
|  | Iteration 5 |  |  |  |
| **Element *P*(*i*)** | **Reachability set: *R*(*Pi*)** | **Antecedent set: *A*(*Pi*)** | **Intersection set: *R*(*Pi*) ∩ *A*(*Pi*)** | **Level** |
| 9 | 9 | 9 | 9 | 5 |